

越努力 越幸运

经管之家自制电子书精彩首发

与计量统计一起 “大冒险”

目 录 I

序：每个人都能学好计量统计

师兄本科几乎不曾接触过统计学、计量经济学，而且在研究生阶段，统计课程大多偏重数理推导，很少得到真正的实践应用，他宛若打了鸡血一样，不断去吸收、去学习、去实践，走出了一条不同于以往——死啃理论基础——的学术之路。从中不难看出这难能可贵又有时代精神的两个品质：摒弃旧有、跨界混搭与从应用场景中学习。

一、计量统计没有想象中那么难：我为什么要写这本书？

计量理论是“道”，是思想；软件实现是“术”，是工具。工具是为解决实际问题服务的，不能本末倒置，不要太注重工具的学习而忽略软件实现背后的理论。我自己的亲身经历大概有个懵懵懂懂——追求计量软件的实现——探索计量背后的理论——理论与软件实现并重这么一个过程。虽然，现在相对于很多科班出身的专业人士来说，我可能还在入门的门口徘徊或者离入门还远，但我始终还是保持一颗虔诚、敬畏的心来学习计量，使用计量。因为学习计量，用计量去解决提出的科学问题是我感兴趣的事，会为之奋斗一生的事。

二、菜鸟如何搞定计量统计：亲历、心得、经验、成败

可能是受长期中国教育模式的熏陶，我一直都比较急功近利，和大多数在人大经济论坛上寻求帮助的人一样，都想快速地学习到我想要的某种计量/统计方法，然后能够利用手里的数据写文章。虽然寒假我将伍德里奇的书扛回家了，然而我最开始也不是去看它——慢慢的，我越看越有劲……为什么写得这么经典，这么好。一些长久困扰我的疑惑，如自变量因变量取对数后模型怎么解释，2 期数据的面板回归怎么做等也慢慢的得到了解决。经过这一系列过程，再回过头来看我这个外行的计量之路，似乎以前走的有点偏。以前是重实用，为模型而模型。现在更想做的是多探究为什么要用这个模型的机理，我想这可能才是一条长久之路。慢慢的现在也意识到了计量与统计的可能不同，计量的博大精深。同时，也发现计量也不是想象中的那么难了……

三、拨开计量与统计的面纱：软件学习精彩答疑与学习资源荟萃

我在写第一篇 SCI 时由于没有系统学习过计量的东西，只是使用 SPSS 来实现自己的想法，这导致文章在实证模型上就弱了一些。我慢慢地发现 SPSS 解决很多问题似乎不够用了，且在书中反复出现 Eviews，STATA 等统计分析软件的字眼。这推着我往前思考，是不是在学习 SPSS 软件的基础上要再另外系统地学习一门统计分析软件。SAS，我在研究生院上课时有过接触，是很强大，然而感觉比较难学，

同时安装起来都感觉比较麻烦。于是我慢慢地缩小范围，决定在Eviews 和 STATA 两个软件里选一种软件学习。

四、走近计量与统计的后花园：多元分析精彩答疑与学习资源荟萃

又遇到一个良心期刊，虽把我拒了，但给了我几页密密麻麻详细的修稿意见，涉及模型的稳健性，变量的内生性问题。那时我才发现以前的心态是多么的可笑，以为看了 SPSS 软件学习培训视频就可以解决以后科研道路上大部分问题，却不知道审稿专家所谓的模型的稳健性和变量的内生性是什么。这让我瞬间有点惶恐，焦灼了一段时间。直到后来各种机缘巧合下慢慢向计量经济学的学习探索之路靠近。

五、采撷计量与统计的蜜果：回归分析精彩答疑 1 与资源集合

我是在各种机缘巧合下开启计量经济学的学习探索之路。仔细想来，实际是解决两个现实问题的迫切需要促使我向这条路靠近。一是正如上文所说的稿件被拒后审稿意见中的模型的稳健性和变量的内生性看不懂的问题，这两个概念已经超过当时我的认知了；二是我自己买的几本专业方面的专著中广泛的涉及到了 FGLS，2sls 等模型，仅仅学了张老师 SPSS 培训视频和看过几本 SPSS 统计书籍的我对这些概念和软件实现更是不懂，看着费力。这两个问题着实困扰了我一段时间，直到机缘巧合下伍德里奇《计量经济学导论（第三版）》的出现。

六、窥探计量与统计的方圆：回归分析精彩答疑 2 与资源集合

我也慢慢发现——计量学习虽重要，但相较而言，文章要解决的关键科学问题可能更重要。前者之所以重要，是因为它是解决关键科学问题的工具，而后者之所以重要，是因为它是文章的灵魂，可能是决定文章被录用与否的关键。有了这个心态的转变，我慢慢地开始反思这一路的计量自学之旅。

七、问在经济学的路上：真正的学习从提问开始

从事设计专业的统计小白被迫走上学术道路，然后面对一堆建模数据和方法选择无从下手，就连最基本的数据描述性分析都不会。误打误撞中来到该论坛的 SPSS 专版提下了一个现在看来很白痴的问题，本来以为会跟其他地方一样没人理睬，但是天降 xddlovejiao 大神热情地对我的问题详细的进行了解释，还推荐了很多专业书籍，让我一步步对 SPSS 和逻辑回归分析有了初步的了解。而且最让人敬佩的是，大神在帖子中不嫌麻烦地针对逻辑回归的原理/数据处理方法等内容，一步步地跟我解释，回答速度也非常快，大有版聊之势

八、为每个经管人解决学习难题：在你身边 为你答疑

知识的价值不在于所有权，而在于它能够让更多人更好地理解这个世界，让人们用更美好的方式来生活。聚焦每个经管人的知识，为每个经管人所用，答疑是学习，也是交流。很多时候，答题与其说发现世界，不如说发现自我。答题是一种经历、知识、智慧的梳理与总结，是个人知识积累的原创公开作品集。我们发起并组织经管答疑，为你，为我，为解决难题，为知识分享，也为智识传递……

九、与大数据共眠：经管之家大数据与数据分析好课不断

如今，我们已经进入了企业发展日新月异的“互联网+”时代----一个用数据说话的时代，也是一个依靠数据竞争的时代。数据的分析

和处理能力正在日益成为企业倚重的技术手段，对数据分析方面人才的需求更加迫切。

附录：其它版主/坛友的读后感和推介

目 录 II

请拿出手机，扫描下方的二维码，阅读更多精彩内容



1. 经济、管理、金融、统计类难题在线解答平台

2. 经管之家丰富学习资源平台





3. CDA 数据分析师学习资源平台

4. 每天一个数据分析师教你学习平台





« 5. 一个外行的计量统计之路

6. ssci/sci 发表的一点心路历程 »





« 7. 北美一流统计学专业课程设置

8. 统计学学习 10 年的经验之谈——致统计学初学者 »





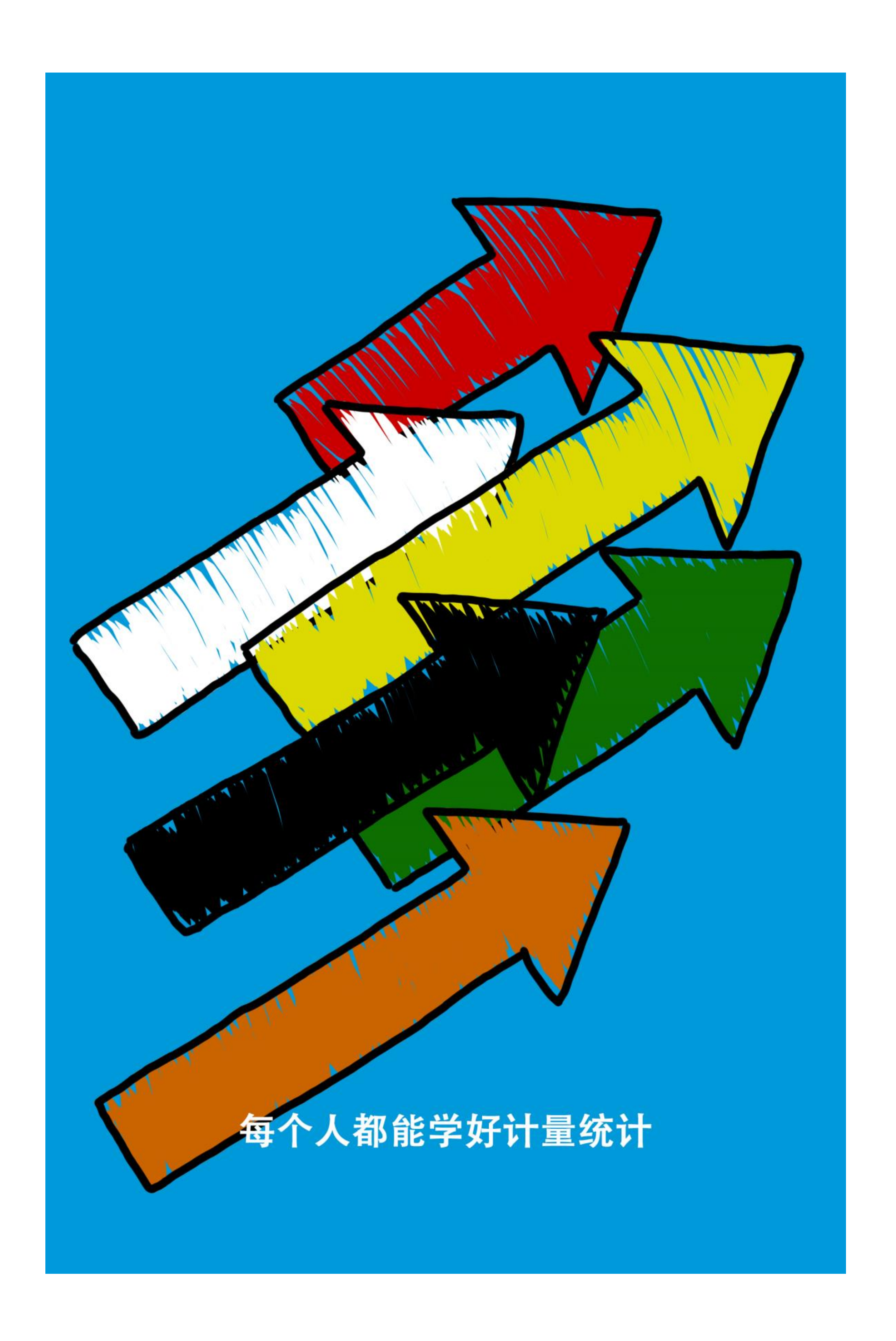
« 9. 【一 VIEW 易 VIEW】一切从回归开始

10. 【时间简“识”】1. 带你看看时间序列的简史 »





« 11. 数据分析师的完整知识结构


每个人都能学好计量统计

序

每个人都能学好计量统计

〖师兄本科几乎不曾接触过统计学、计量经济学，而且在研究生阶段，统计课程大多偏重数理推导，很少得到真正的实践应用，他宛若打了鸡血一样，不断去吸收、去学习、去实践，走出了一条不同于以往——死啃理论基础——的学术之路。从中不难看出这难能可贵又有时代精神的两个品质：摒弃旧有、跨界混搭与从应用场景中学习。〗

序一：

Think bigger！本书诞生记！打开它的正确方式

文｜经管之家经管答疑团队

互联网已成江湖，很多人迷恋、沉浸。唯独就是不爱电子书。

在网上读书的用户体验是很牵强的，纸质书的质感和香气更像是真爱。有时候，我们自己都觉得，把一本已经在网上被人们熟悉的文字转成电子版供人阅读，这个动作本身就像是个圈套。但是，那细碎积累和思考意念下的价值，让我们想通了：不管电子书还是纸质书，好书才是硬道理。于是，您就看到了这本小小的书。

这本书，我们赋予它一个主题，名曰“一个外行的计量统计之路”。其实，我们是故意的。一般来说，小伙伴们更爱看行家、专家、大牛、前辈们的作品，更权威、更精进、更高大上。但我们琢磨着，对充满学习欲望的求知者来说，最了解初学者的永远是初学者，您踩过的坑，跳过的陷阱、撞到的篱笆，别人都遇到过。如果，您又恰好正是一位有计量统计学习需求的人士，有人用一本书告诉您避雷大法，您会不会觉得很有用？很有趣？很贴心？甚至于一些轻微的窃喜、莫

名的快乐和冷暖互知的得意。

同样，前车之鉴，后事之师。在今天的大数据时代和 IT 时代，计量经济学与统计学的重要性众所周知，然而学好它所要经历的痛点、难点、苦点，也往往戳中大家的泪点。试想，如果有人把学习计量统计的真实记录、心得感受、得失成败，毫无保留地传递给您------甚至是以莫大的信心和勇气对计量统计学习过程进行透彻的解剖，而这过程就像自己拿着刀为自己开刀，那么我们想，这就绝不是轻描淡写、不痛不痒的事儿。这更像是，在计量统计的深水河中，被静静的打磨着的那些精美的鹅卵石。Think bigger 的意念一闪，我们决定将它重新编辑、加工、补充、修订，最终整理为一本书。

这本书是一叠原创的碎片。但它是有序的，而非随意的。

这本书有为数不少的、长文型、梳理型的原创，它们大多很丰满，言之有物。也是作者若干年来学习和求索计量统计的真实记录。其实读原创部分的时候，您可以随机跳跃。 但每一篇原创都是学习的小矩阵，内在的逻辑是严密的，分享是真实的，思考是独到的。

这本书是一本计量统计答疑解惑的学习手册，但你也可以当成搜索引擎。

异方差、多重共线性、统计报表、数据分析、多元分析、面板数据、特征值、T 分布、显著性检验、共线性消除、非参数检验、自变量、中间变量、因变量。如果这些难题曾经让您迷茫或切齿。如今您可以搜来翻去，各个击破。这本书也

是经管之家组织经管答疑的产物，本书作者正是我们邀请的答疑专家之一。他的回答，认真负责，条分缕析，直中要害。更难能可贵的是，透着一股浓浓的、化不开的认真。而认真，是我们最为欣赏的品质之一。

这本书能找到 100 多个计量统计经典难题的答案。但你还是可以继续向我们提问。

如果您在书中，把您的部分未知，通过答疑变成了已知。那么恭喜您，还可以继续通过经管之家经管爱问平台，或微信号 jgasker 向我们提问。我们组织专人答疑，随时丢您的难题过来，随时找到答案。

Think bigger！这本书也是经管之家答疑团队与您的赤诚相见，但我们绝不裸奔。

我们相信，总是从更大的层面思考问题，是有益处的，一如这本书从文章到篇章再到一本书。Think bigger！无论是啥，当您用一百倍的尺度思考了某个问题之后。您会发现很多东西的真面目，拉长时间的维度，思考的放大，坚持下来，会是什么局面？有些东西已经放大，bigger 之下，我们与您以书会友。但生活还在继续，形而上，形于内，理性和坚持的力量同在。

最后，这本书是一部对学习和求知无始无终的穿越。不说再见。如果您想了解我们的团队，还有我们，可以戳这本书的第九部分。

经管之家经管答疑团队
2015 年 12 月

序二：

像外行一样思考 像专家一样实践
每个人都有可能学好计量与统计

文 | 郑玉平 全国大学生统计建模大赛一等奖获奖者，经管之家计量统计区版主

畅销书《像外行一样思考，像专家一样实践》提到：像外行一样思考，可以给我们完全不同的视野。很多时候，换个角度看问题，问题往往就很容易想到新的解决方法。

我觉得这个理念用在本书作者——定德师兄(下称师兄)身上再恰当不过了。师兄本科几乎不曾接触过统计学、计量经济学，而且在研究生阶段，统计课程大多偏重数理推导，很少得到真正的实践应用。师兄抱着好奇心与求知欲，顶着“计量零基础”开始学习计量经济学与张文彤老师的 SPSS 系列丛书，后来在写论文的时候需要用到分位数回归，SPSS 已经解决不了，于是又转战 stata 软件的学习。随着理论、软件学习的深入，让人拍案叫绝的案例为师兄打开了新的天地，他宛若打了鸡血一样，不断去吸收、去学习、去实践，走出了一条不同于以往——死啃理论基础——的学术之路。

从中不难看出这难能可贵又有时代精神的两个品质：摒弃旧有、跨界混搭与从应用场景中学习。

第一点：摒弃旧有、跨界混搭。木心先生在《从前慢》就专一提到："从前的日色变得慢；车马邮件都慢；一生只够爱一个人。"从前的学术就是这么"绝对痴心"，买断该专业所有书，放在家里备上尿壶，除了出门觅食，就是闭关看书。学富五车之后，您就是专业大拿，摆好姿势，跟人合影，十块一张都够活了。

现在却不同，连小学辅导书都出了千千万万，在专业极度细分与相关书籍枝繁叶茂的情况下，您要穷尽所有专业书籍，还真不是那么容易的事。或许，通过跨界混搭能开出不一样的奇异果。

如今，传统企业嚷嚷着互联网转型；互联网企业会去做电影；演员与歌手又会去做创新企业投资。数据科学家也是统计学家与计算机科学家之间的混搭。师兄农业经济百搭计量经济，出现了如今专属他的春天。

第二点：为自己搭建学习的应用场景。高校教育多推崇数理推导，认为理论扎实了再扔到社会去学应用即可。这样的思路无可厚非，但这并非最优化的学习逻辑，我们需要看到，实际的社会企业应用环节中，理论的需求往往位于较为次要的地位。入门数据分析用 SPSS、JMP、excel 等软件就可以直接实现，并不用编译内置代码，这些图形菜单驱动界面的软件极大方便新手使用；对于深一些用 stata、R、python 等编码，很多内置的软件程序包被无数次科研实践过，已经

久经考验，只需要简单地修改，就能适应复杂多变的现实情况。

谈及应用场景，导师科研、学术竞赛、理论博客等地方都会提供很多、很优质的案例以让您实战模拟分析。所以，把理论还给应用，从社会实践中学习理论，已成为不可逆转的大趋势。此时，如同本书一样，将应用的理念传达出来，显得尤为重要。

师兄在学术上是严谨、睚眦必较的小学究形象，私下里却是一个欢脱的“二货”。百搭人生，百搭学术生涯，充分符合“有混搭胆子，就这么任性”的宗旨。另外，师兄在经管之家（原人大经济论坛）一腔热心肠、有问必答，解决很多坛友的学术问题，竟然还有匿名粉丝怒送大闸蟹，真让尔等心生羡意。

最后，衷心希望师兄的热心肠与百搭的经历能给大家一点启发与鼓励，像外行一样思考，像专家一样实践，找寻专属您的春天。

郑玉平

序三：

计量虐我千百遍 我待计量如初恋

文｜杨冬 西南财经大学博士在读，经管之家计量统计区版主

计量经济学在中国的发展到今年恰好是 35 个年头，1980 年在北京颐和园举办的计量经济学讲习班对于促进这一学科在国内的发展无疑是深远的。时至今日，很多计量经济学著作都被国内汉化出版，这对于像我这样所谓的土鳖博士来说，也算是省去了好多阅读上的时间。

算起来，自己从接触这门学科至今也已经有 6 年的时间，出于对这门学科的喜爱，自己也从统计学的学士、硕士转到了数量经济学的博士，并进一步从事时间序列分析相关领域的学习。

计量经济学这门学科虽然和其他比起来比较年轻，但它的魅力通过近些年经济学、社会学等方面的学术论文传播便可见一斑。这也使得很多同学都想迅速地踏入计量经济学的门槛，更好地使用实证研究方法，而经管之家（原人大经济论坛）推出的这本《计量统计历险记》也就是在这样的大环境下逐渐成形，它的目的就是给大家一个更快、更好的方法，了解和学会必要的基本和中等的计量技术，

从而更好地进行实证研究。同时这本书也为大家展示了作者以一个“外行”身份，学习和摸索计量统计规律、掌握方法、求知推进的过程，有很多的亲身经验和独特的思考，相信能给想一窥计量/统计神秘面纱的您一些启示。

一般来说，我们习惯把数量经济学研究领域分为四大类：时间序列分析、面板数据分析、微观计量分析以及非参和半参数计量分析。当然，这只是个简单的分类，现代计量经济的发展是迅速的，Ling(2015)在 JOE 上的论文归纳了近些年来计量经济学发展的 20 个前沿方向，很多研究已经不在局限于上述的分类。自己在参与汇总电子书的过程中大概也是按照这个思路提供了一些素材和资料。

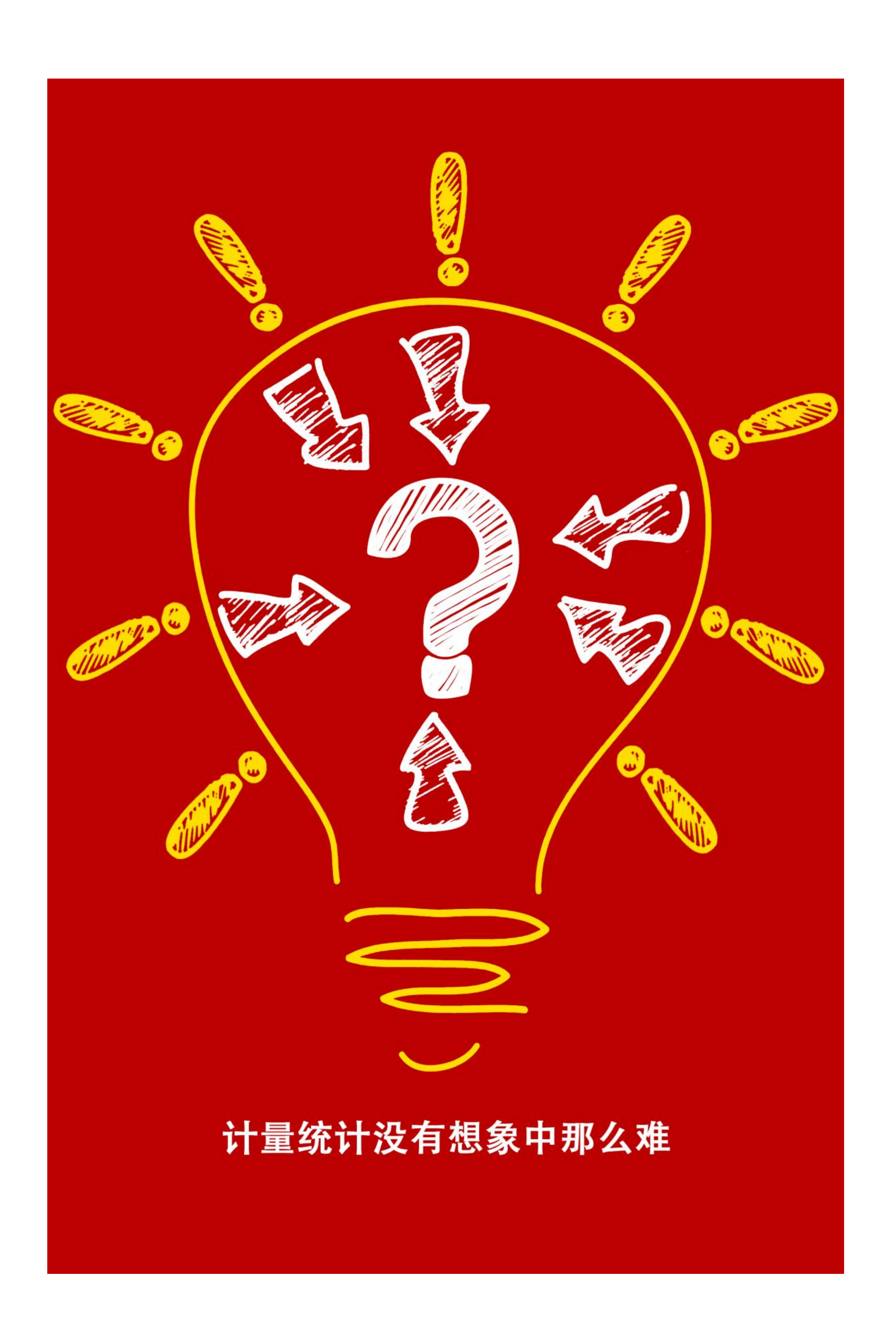
很多人也都在问我，如何才能学好计量经济学这门工具，我想这个应该是因人而异吧，对于不同专业，不同目的，对计量认识的深度自然也应该有所不同，如果是非经济学专业的学生，在时间有限的情况下，个人觉得通过实践去加深对计量的认识可能会更快地了解这门学科。而对于经济学专业非数量的学生，在掌握基本的计量方法上，可以结合部分书籍和论文去尝试写作。当然，这个只是自己的一些心得和体会，并不是所有人都会适合。对于软件的使用方面，论坛里可以说已经基本涵盖了所有主流的统计和计量软件，从使用的频率来看，R、STATA、Eviews、SPSS 相对较多一些。我一直觉得，软件没有什么好坏之分，不能说 R 就一定是最好的，也不能说 Eviews 就是最低级的，软件作为一个载体不过是实现你研究目的的工具，如果熟悉和掌握了任何一门软件的语言，都可以让您很好

的解决大部分问题。

在计量经济学书籍的选择方面，如果想从基础开始学习，格林、伍德里奇或者古扎拉蒂的书籍无疑是最好的选择，看完其中任何一本著作都会令大家拥有巨大的收获，当然，也可以结合软件书籍学习，这样会达到事半功倍的效果。

在我看来，计量经济学是一位年轻而又美丽的姑娘，有些人说她华而不实，有些人抱怨看不懂她的“心”。方法本无对错，它是我们认识世界的一个工具，如果不是为了使用而使用，那么计量经济学无疑会给您更丰硕的结果！

杨冬


计量统计没有想象中那么难

一、

计量统计没有想象中那么难：我为什么要写这本书？

〖计量理论是“道”，是思想；软件实现是“术”，是工具。工具是为解决实际问题服务的，不能本末倒置，不要太注重工具的学习而忽略软件实现背后的理论。我自己的亲身经历大概有个懵懵懂懂——追求计量软件的实现——探索计量背后的理论——理论与软件实现并重这么一个过程。虽然，现在相对于很多科班出身的专业人士来说，我可能还在入门的门口徘徊或者离入门还远，但我始终还是保持一颗虔诚、敬畏的心来学习计量，使用计量。因为学习计量，用计量去解决提出的科学问题是我感兴趣的事，会为之奋斗一生的事。〗

1.1
我是谁？我写什么？

引子

人常道，有心栽花花不开，无心插柳柳成荫。

其实，从没想过，我也可以用一本电子小书的形式与您谋面。

几个月前，突然接到经管之家（原人大经济论坛）编辑（同时也是我师组）的消息，说她有一个想法：想将我的计量实证自学之路的故事和心得，以及我发表过的原创文章，我在计量统计领域回答坛友的若干经典疑难问题，编辑为一本电子书，希望透过我的学习历程给更多计量统计初学者以信心和启发。乍听之下，我内心既激动又忐忑------激动的是没想到我人生中的第一本书会以这种形式、这种内容来和大家见面，这让我这个中科院的小博士有点小小的意外，“脑袋”一热就答应师姐尝试一下。事后细想，突然感觉比较忐忑。经管之家（原人大经济论坛）近 700 万会员，卧虎藏龙，多的是比我厉害得多的“大牛”，我一个计量自学出道的外行有什么资格，有什么能力在这些“大牛”面前班门弄斧呢？

后来把担忧和师姐一说，师姐就劝我："书籍是传播的最好形式，有价值的东西就值得被人们知道。虽然是以电子书的形式和大家见面，但书的格式、样式和排版制作等都会参照正规出版物流程，最重要的是，只要书的内容对大家有启发和意义，对得起读者的阅读时间，传播出去又有何妨。" 师姐的鼓励打消了我很多顾虑。于是乎，在论坛团队的协助下，前后经过几个月的努力，有了您面前的这本小小的电子书。

说到书，自然少不了作者简介、写作目的之类。师姐让我也写一个，并建议说："言为心声，文字不拘表象，只要真实；表达不必严肃，只要言之有物"。经过一段时间的思考，暂时没理出个具体的头绪，故而一拖再拖。后来，在论坛答疑团队的帮助下，书的大纲被拟出来了，看着大纲，我灵光一闪，想着可能将我与经管之家的缘分，我在学习计量统计时每个重要时点和阶段所进行的思考总结，一起真实地展示出来，可能对梳理全书的脉络能起到不少帮助，也能让感兴趣的读者看到我在计量统计学习路上，不同阶段的心路变化，以及在哪些情况和机缘下发生的变化。于是，这本书渐渐清晰了起来……

第一次与经管之家（原人大经济论坛）亲密接触

我于 2013 年在中科院研究生院上课时因为要下载统计年鉴而注册人大经济论坛（现已改名为经管之家），与论坛有了第一次接触。然而，当时只是为下载资料而注册论坛，没有进一步意识到论坛的价值。故而下载资料后就没有过多地关

注论坛了。直到 2014 年 8 月份和一个高校的同门师兄去凉山州出差，偶然间的交流知道国内 SPSS 教材畅销书作者张文彤老师在人大经济论坛办有 SPSS 视频培训班。我对张老师膜拜已久，原因有三：

一是在研究生院上《心理多元统计》这门课时，老师推荐的辅助教材就是张老师的《SPSS 统计分析教程》，我自己也在网上买了本仔细地拜读了一下，感觉写得很好；二是我自己学习之余，尝试写的小文章用到的有序多分类 logistic 回归模型就是在此书的指导下完成的，尝到了点甜头；三是我从本科开始接触 SPSS，陆陆续续上过 3 门会用到 SPSS 的课，然而几门课的老师基本都是做自然科学实验的（如有做农业实验的，有做氮循环的），所以所学内容基本都在 t 检验，方差分析和简单的多元线性回归里绕，始终感觉学的不系统，不扎实，知其然而不知其所以然。例如具体在做各个检验时，为什么要勾选一些对话框，而不勾选另外一些，每个对话框的设置有何用基本都不知道，只是在老师的指导下“傻瓜式”的重复操作过程。

所以机缘巧合下，得知网上可以买到张文彤老师的 SPSS 培训视频，出差回来后我做的第一件事就是找视频，然后如痴如醉的将其系统的学习了一遍。在学习视频的同时，在网上了解到张老师出了一本新书《IBM SPSS 数据分析与挖掘实战案例精粹》，也有配套视频。于是乎，配合着张老师的书和视频，我的 SPSS 技能得以快速的提升。俗话说光掌握理论不行，还得实战，在实战中发现问题，

进一步查缺补漏。于是乎，我再次来到人大经济论坛，活跃在 SPSS 版块，帮助坛友解答一些基础问题的同时，也进一步巩固自己的 SPSS 知识。结合看视频的心得和自己的一点思考，我想应该写点什么。故而有了“SPSS 数据分析心得小结”这篇试水贴。帖子发出后，得到很多坛友的肯定和支持。这让我倍受鼓舞，投入更多激情到 SPSS 软件和书籍的专研中去。这算是注册后真正意义上与论坛的第一次亲密接触。此后缘分越结越深，一发不可收拾。

SPSS 学习后的牛刀小试：写作缘起

身为学生，面临着毕业和将来找工作等各方面的压力，所以在北京研究生院学习期间我就尝试着利用导师项目的调研数据写了两篇小文章。一篇是纯粹的描述性统计分析文章，一篇是我上文中提到的模仿着张文彤老师《SPSS 统计分析教程》中的有序多分类 Logistic 回归模型写的文章。由于是新人还没入门，加之还没有系统地学过 SPSS 或其它统计软件，所以这两篇文章也算是饱受挫折，被拒多次（心得详见“农林经济类文章发表个人经验——两个学报和一个农业科学期刊”）。

结束研究生院的课程学习后回所，机缘巧合下学习到了张文彤老师的 SPSS 培训视频，统计建模能力有所提升。幸运的是同时中国农业大学学报退修，我利用视频中学到的知识重新对模型进行了构建和解释。经过一些挫折后，文章得以顺利发表。后来又趁热打铁，又写了一篇中文文章，投稿到西南大学学报去了，

最终也得以顺利被收录。

虽然现在回过头来看，自己的前几篇文章还存在这样或那样的不足，甚至生平的第一篇文章在最新版的北大中文核心中被剔除了，但人生难得的就是从学术菜鸟到学术基本入门的这个奋斗过程。尤其是在没有人手把手指导，独自“享受”文章写出来不知质量怎样，一投就被拒，被拒又再投，再投又被拒，如此反复直到有好的期刊给予建设性的意见，照着意见修改后文章被收录为止。一路奋斗，一路硬着头皮“闯”的艰辛可能只有有过类似经历的人才能体会。现在想想，奋斗的过程又何尝不是享受的过程呢。

那时的我有了几篇小文章变得有点飘飘然，以为学了张文彤老师的 SPSS 培训视频就可以解决我以后科研道路上的大部分问题，以为统计就是计量，没啥高深的（当时心态的具体描述详见“SSCI/SCI 发表的一点心路历程”）。也正是这种不敬畏不虔诚的心态导致了后面 SCI/SSCI 投稿的多次悲剧，开启了“碰壁”之旅。直到多次被拒后，又遇到一个良心期刊，虽把我拒了，但给了我几页密密麻麻详细的修稿意见（涉及模型的稳健性，变量的内生性问题）。那时的我才发现以前的心态是多么的可笑，以为看了 SPSS 培训视频就可以解决以后科研道路上的大部分问题的我却不知道审稿专家所谓的模型的稳健性和变量的内生性是什么。这让我瞬间有点惶恐，焦灼了一段时间。直到后来各种机缘巧合下慢慢地向计量经济学的学习探索之路靠近。软件由 SPSS 慢慢向 stata 过渡，心态上也由飘飘

然变得更加敬畏和虔诚。

后来一路摸索，就这样参加了人大经济论坛组织的经管答疑，再加上之前就计量与统计版块写了若干篇学习原创的经验分享，于是，机缘巧合之下，就有了摆在您面前的这本小书。

1.2

一个外行的计量经济学学习探索之路

诚如上文所说，我是在各种机缘巧合下开启计量经济学的学习探索之路。仔细想来，实际是解决两个现实问题的迫切需要促使我向这条路靠近。一是正如上文所说的稿件被拒后审稿意见中的模型的稳健性和变量的内生性等看不懂的问题。这两个概念已经超过当时我的认知了（即使找遍了张文彤老师的 SPSS 培训视频也没发现相关字眼），百度搜索要么是一些概念性的介绍（相对比较泛的东西），要么就是一堆公式在那里，看着我头大；二是我自己买的几本专业方面的专著中广泛的涉及到了 FGLS，2sls 等模型，仅仅学了张老师 SPSS 培训视频和看过几本 SPSS 统计书籍的我对这些概念和软件实现更是不懂，看着费力。这两个问题着实困扰了我一段时间，直到机缘巧合下伍德里奇《计量经济学导论（第三版）》的出现。

机缘巧合下我发现了学姐办公室中的《计量经济学导论（第三版）》，看到里面有关于 2sls 的描述，如获至宝。当时又恰逢寒假，于是乎不辞辛苦地用皮箱把

两本大块头的书拖回去，硬着头皮把横截面数据分析部分啃了两遍，受益匪浅(过程详见“一个外行的计量经济学学习之路”)。

在学习伍德里奇书的同时，我慢慢的发现 SPSS 解决很多问题似乎不够用了，且在书中反复出现 Eviews，Stata 等统计分析软件的字眼。这推着我往前思考，是不是在学习 SPSS 软件的基础上要再另外系统地学习一门统计分析软件。Sas 我在研究生院上课时有过接触，是很强大，然而感觉比较难学，同时安装起来都感觉比较麻烦。于是我慢慢缩小范围，决定在 Eviews 和 Stata 两个软件里选一门软件学习。

有了这个想法，我到论坛相应的版块去搜索。因为前面学习 SPSS 时尝到了视频学习的甜头，所以我搜索的重点从两门软件的视频培训入手。同时，也搜索资料比较了两门软件各自的优劣。说来也巧，我一到 stata 板块搜索，就看到传说中的 stata 大牛连玉君老师的很多帖子了。同时，我也发现 stata 版块的活跃度要比 Eviews 要高一些，且 stata 在特性上更加短小精悍，在回归这块有其特殊的优势。综合比较下我决定学习 stata 软件。

要重新学习一门软件，谈何容易。第一件事就是找准适合自己的资料。有了 SPSS 的学习经验，我通过各种渠道搜索，最终购买到连玉君老师的培训视频。同时，搜索坛友公认的学习 Stata 比较好的书，如现在依然放在桌上的工具书——陈强老师的《高级计量经济学及 Stata 应用》，汉密尔顿《应用 Stata 做统计分

析》等，第一时间把纸质版的买回来，边学习视频边看书。慢慢的开始上手使用 Stata，并参考连玉君老师的培训资料一步步把自己第二篇 SCI 的模型构建了一遍。

在学习的同时，我也经常关注人大经济论坛 SPSS 和 Stata 版块，发现一个现象：基础问题提问者居多，分享一些“高大上”的资料（各种统计分析资料的英文版）的人居多，而实实在在的分享一些计量/统计学习心得的人少。恰逢我自学计量一段时间，感觉应该写点什么回顾下我这一路的心得历程。于是乎我斗胆之下写了“一个外行的计量经济学学习之路” 这篇帖子，没想到一下子就引起了很多人的共鸣，很多坛友发帖回复表示肯定、鼓励和支持。回帖量一高，慢慢地引起了论坛 HR（观世鹰版主）的注意，开始了我和论坛更深的缘分——邀请我来做论坛经管代码库的版主。天性好奇的我觉得新鲜，决定尝试一下。同时，发挥干中学的精神，边学习边实践。在快速地过完一遍连老师的视频后，我尝试着自己用 Stata 软件来实现自己论文的模型。进而有了下面的一个外行的计量经济学实证之路。

1.3

一个外行的计量经济学实证之路

正如我在前文所说 我在写第一篇 SCI 时由于没有系统地学习过计量的东西，只是使用 SPSS 来实现自己的想法，这导致文章在实证模型上就弱了一些。同时，由于文章写作我也没有受过系统的训练 ,故而我的写作习惯/行文思路可能也和外国人的不一样（SCI/SSCI 写作其实有一定的范式/套路），再加上其它的一些原因使得我第一篇文章多次被拒。当然，以上的一些问题也是我事后才想明白的，当时是没想明白的，以为中了几篇中文，英文按照中文写作的套路去也没问题，只是一个是中文表述，一个换成英文表述去了，而实际上还是有很大差别的。

虽然我的文章接连被拒，然而幸运的是，我在学习伍德里奇《计量经济学导论》时，发现里面有一章专门讲实证文章写作思路的，对于那时的我简直是醍醐灌顶。那时我才突然意识到为什么我的文章会接连被拒了（“SSCI/SCI 发表的一点心路历程”，这篇帖子里有详细描述）。同时，我的第一篇文章遇到了一个良心期刊，虽然把我拒了，但给我提了 4 页密密麻麻的修稿意见。这些意见中对应着

我文章致命的东西要恰好和伍德里奇书籍中讲实证章节部分的写作思路对应。这让我进一步加深了对英文文章写作套路的理解。于是乎，我痛定思痛，大修甚至重写了第一篇文章，使其结构基本符合英文稿件范式，并利用这段时间计量所学所想重新构建了模型。最终再投出去，耗时几个月文章得以被顺利发表。

有了前面血淋淋的经验，我第二篇 SCI/SSCI 文章就要顺利得多（当然，这里面还得感谢国外一个专业的老师帮我文字润色和提出的宝贵修改意见），最终经历一次大修，一次小修得以顺利发表（心得详见“SSCI/SCI 发表的一点心路历程”，这篇帖子里有详细描述）。

有了两篇英文文章发表的经历，我知道我的英文写作之路可能慢慢的已经入门，以后再按照这个套路去写，可能会被拒，但至少应该不会再像第一篇文章那么惨。同时，我也慢慢发现——计量学习虽重要，但相较而言，文章要解决的关键科学问题可能更重要。前者之所以重要，是因为它是解决关键科学问题的工具，而后者之所以重要，是因为它是文章的灵魂，可能是决定文章被录用与否的关键。有了这个心态的转变，我慢慢的开始反思这一路的计量自学之旅。

1.4
一个外行的计量经济学热心答疑之路

（注：本文中提到的答疑是指，2015 年起，人大经济论坛组织各领域专家和热心网友，集中解答经济、管理、金融、统计软件与数据分析等经济管理类学习难题，本书作者也被邀请回答网友们提出计量统计的难题）

在两篇 SCI/SSCI 写作、投稿、修稿和收录这期间，因为“一个外行的计量经济学学习之路”这篇帖子，我被观世鹰版主邀请担任论坛经管代码库版主。版主实习期为两个月，两个月间需要举办一次版面活动。我仔细的思考了一下，结合自己的计量实证历程写了“’万万’个论坛币计量实证的’摸底’考试”这篇帖子。由于目前发的文章全部是基于横截面数据的，所以这个帖子的内容基本也是针对横截面数据（面板数据处理的心得等以后发过相关文章，有更进一步的学习思考后再补充）。从变量类型的判定、变量的预处理、方法的选择到结果的解释，基本是遵循实证的逻辑。这个帖子出来后引起了很多坛友的兴趣，很多热心的坛友也结合自己的经历来尝试的回答了其中的一些问题，也有一些坛友问我是否有

这些问题的个人看法（类似于答案的东西）。非常抱歉的是这个帖子开了后，我一直被各种杂事缠身，没有时间单开一贴来系统的分享下我对这些问题的心得，前期只是尝试性的写了一篇极端异常值的诊断帖子（“描述性统计分析在实证研究中的作用及具体软件实现——以 SPSS 为例”）。幸运的是现在帖子的回复中已经有很多坛友的优秀见解。同时，在本书后文的精选答疑部分，很多帖子的回复都涉及到此贴中的问题，也算是对此贴的一个有力补充。

在担任版主的同时，考虑到经管代码库才成立没多久，除了论坛的版主外，可能很少有人知道论坛还有这么一个版块。故而，为了增加版块的人气，我结合自己的 SSCI/SCI 写作心路历程，写了一篇名为“SSCI/SCI 发表的一点心路历程”的帖子，发在学术道德监督版块。没想到的是，帖子发出后又得到了很多坛友的肯定、鼓励和支持，并引起了论坛编辑团队师姐的注意。这个时候，恰逢论坛大力组织经管答疑版块，也就是后来的经管爱问。于是乎，师姐找到我，问我是否愿意加入专家答疑团队，帮助坛友回答一些计量统计上的问题。起初我本很忐忑，因为相比于计量统计出身的专业人士而言，我始终是一个外行，是靠自学出家的。担心自己自学的能力有限，不能精确有效地回答坛友来询问的一些问题。最终在柴柴起师姐和我的素质低师弟的鼓励下，决定尝试一下，就这样，开始了我与论坛的经管答疑之缘。同样没想到的是，仅仅过了几个月时间，经管爱问便成了广大经管学生和研习者，提问和讨论的知识交流重地，每天都有成百个难题在此出

现，等待回复。

可能因为我是以一个“自学外行”的身份来参与经管答疑，所以我多少能体会到来论坛提问和求助者的心情（在某种程度上我自己前期也是这么过来的，我在一个外行的计量经济学学习之路中有具体描述）。所以当时只要我会的，或者我知道在什么地方能找到和这些问题相关的资料的问题，我都会第一时间给予解答。有时感觉在论坛的帖子里或者消息里解释不清楚，就在坛友的同意下远程操作进行指导。然而，同时我也明白授人以鱼不如授人以渔，故而即使远程指导，我也尽量给予大步骤的指导，以及背后为什么要这么做的原理上的阐释，具体的实现过程由坛友自己去摸索，遇到问题了再进一步交流探索。于我自己而言，也算一个学习、巩固和提升的过程。可能也正是因为我的这一份热情，得到了很多坛友的认可和鼓励（甚至还有坛友为了感谢我给我邮寄了好吃的零食，有的匿名给我寄了阳澄湖大闸蟹，在此表示感谢），打消了我作为一个计量外行的疑虑，也让我自信了不少。

答疑半个月左右，论坛就做出战略性调整，由原来的人大经济论坛改名为现在的经管之家。可能由于改版时恰逢暑假，提问开始减少。在这种情形下我还是一如既往的答疑。就这样大概答疑两个半月左右，通过答疑的几百个问题我慢慢发现这些问题背后可能存在某种共性的东西，萌生了再写一篇有助于帮助计量统计提问者的文章的想法。后来有一天在 stata 版块回答一个坛友的提问时，和夏

目贵志版主关于有些基础可能没掌握好的坛友，直接来提问高深计量模型的问题是否需要回答的问题进行了讨论，这促使我进一步反思计量学习的目的和答疑的目的，于是“一个外行的计量经济学学习之路二——经管答疑有感”这篇帖子诞生了。

1.5

一个外行的计量经济学学习反思之路

如前文所说，在梳理 SCI/SSCI 发表心得时，我学习计量的心理就有所转变，慢慢地开始反思、自省。到后面参与论坛答疑两个半月，发现一些答疑共性的东西，再到和夏目贵志版主关于答疑问题的讨论。在这些因素的推动下，我利用一晚上时间写了“一个外行的计量经济学学习之路二——经管答疑有感”这篇帖子。

帖子可分为两部分，前半部分是后期学习计量的一点心得体会，第二部分是答疑的一点体会。帖子发表后，又引起了很多坛友的共鸣和讨论，并被很多网络课堂的老师所收录，这点是我写这个帖子前没有预料到的。

现在再回过头来看我个人的计量自学之路，如果能给渴望一窥计量统计面纱、却又不知何处入手的人，或是给有计量基础但又苦于写不出文章以及文章老是遇坎的人一些启示，那么这本电子书的目的就达到了。回首来路，我感悟到，计量理论是“道”，是思想；软件实现是“术”，是工具。工具是为解决实际问题服务的，不能本末倒置，不要太注重工具的学习而忽略软件实现背后的理论。我自己

的亲身经历大概是这样：最初懵懵懂懂——追求计量软件的实现——探索计量背后的理论——理论与软件实践并重，循环往复的过程。虽然，现在相对于很多科班出身的专业人士来说，我可能还在入门的门口徘徊或者离入门还远，但我始终还是保持一颗虔诚、敬畏的心来学习计量，使用计量。因为学习计量，用计量去解决提出的科学问题是我感兴趣的事，会为之奋斗一生的事。

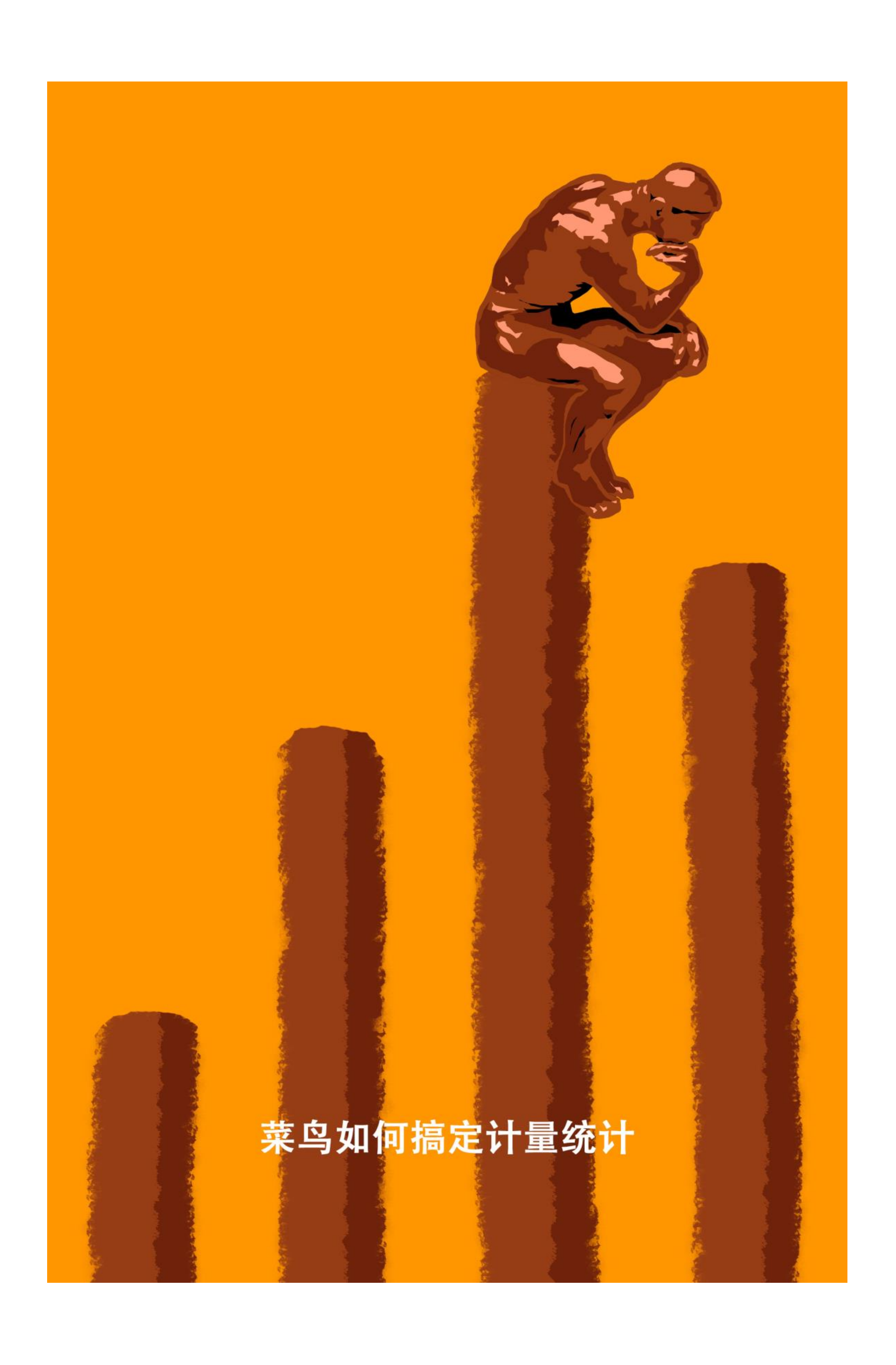
然而，可能您和我都要切记的是：在注重计量理论和工具实现的同时，如果真心想搞学术，还得坐的住冷板凳，要多看文献，多思、多想、多写。我始终坚信：越努力，越幸运，与君共勉。

最后，这本电子书的出现要感谢很多人。感谢人大经济论坛的 admin 对电子书的重视，组织一个大的团队帮忙运作整本电子书；感谢“柴柴起“师姐的赏识和提携，感谢她将出电子书的创意想出来并不留余力的协调团队去实现它，感谢她对整个文本结构的安排和对文字的进一步修饰；感谢以“胖胖小龟宝”、“Rolly”、“我的素质低”和“crystal8832”、”yangyuzhou”、“客初”等为代表的论坛答疑团队的老师和版主同僚，谢谢你们对我答疑问题的梳理，答疑问题前基础知识的整理介绍和对本书其它方面的帮助；感谢为本书付出辛勤劳动的版面设计编辑和其他工作人员；感谢在论坛中给予我帮助过的所有热心版主和坛友；感谢课题组老师、师兄师姐、师弟师妹及家人给予的鼓励；最后，感谢我亦父亦师亦友的导师刘老师，感谢给我提供平台让我尽情地去摸索，在我迷惑的时候，

在我懈怠的时候，在我骄傲放纵的时候给予的鼓励、指导和教导。这本书的面世是大家集体的智慧，与大家合作和共事是我的幸运，再次感谢大家。

xddlovejiao1314

2015 年 12 月 3 日于蓉城


菜鸟如何搞定计量统计

二、

菜鸟如何搞定计量统计：亲历、心得、经验、成败

〖可能是受长期中国教育模式的熏陶，我一直都比较急功近利，和大多数在人大经济论坛上寻求帮助的人一样，都想快速地学习到我想要的某种计量/统计方法，然后能够利用手里的数据写文章。虽然寒假我将伍德里奇的书扛回家了，然而我最开始也不是去看它——慢慢的，我越看越有劲……〗

2.1

初入计量之门：系统回顾我的学习摸爬之路

原文地址：http://bbs.pinggu.org/thread-3596375-1-1.html

原文名称：一个外行的计量经济学学习之路

作为一个非计量非经济学专业出身的人（学过一些统计，SPSS 和 STATA 基本操作），很想梳理下自己学习计量这一路的历程，或许对于初入计量之门，或像我一样非计量专业出身但又想很好地利用计量工具的人，有一些帮助和启示。

第一阶段：计量就是统计——很难很高深

我算是农林经济管理出身，学过简单的 SPSS，那时教我 SPSS 的老师是搞农业种植的，故而传授的多是关于 t 检验，方差分析，简单的回归分析，数据的描述性统计分析等方面的内容。由于大学期间不用发文章，毕业论文也是做的生态方面的东西，加之大学老师只讲操作，不讲原理，所以大学学的 SPSS 基本算是还给老师了。那个时候一度觉得 SPSS 是很高级的东西------因为不知道软件操作

背后原理，也不知道怎么解读结果。

后来有幸被保送到中科院系统一个研究所开始研究生阶段的学习。中科院系统的学生，有一个好处就是会在北京研究生院集中学习一年。由于专业研究的需要（跟着导师做农户微观实证研究）以及自己特别感兴趣（想一窥统计的神秘面纱），故而在北京学习期间选了 6 门关于统计的课（如《心理多元统计》、《统计分析与 SAS》实现等）。其中有些课老师讲得很好（如胡良平老师讲的 SAS，通俗易懂），但可能限于学时，老师没有铺开来讲。加之那时还有其它专业课需要学习，故而除了课堂听讲、完成作业和课程报告外，没有过多的时间去进一步看书消化。经过这么一轮训练下来，一些具体的软件实现大概知道了，为什么这样做也知道一点，但理解还不透彻，结果的解读似乎也是半知半懂。那时候，我的统计知识体系算是比较凌乱、不系统，也一度以为统计就是计量，以为已经揭开了它的神秘面纱了，然而事实却并非如此。

第二阶段：学习统计/计量还得从实践中来

正如我在论坛上发的个人投稿经验交流贴中所述，和大多数在论坛上寻求帮助的人一样，都想快速学习到我想要的某种计量/统计方法，然后能够利用手里的数据写文章。那时候多少有些功利心作祟，有点为写而写——可能大多数人也会经历这个阶段。但是问题来了，虽然我学了那么多门统计课，然而由于教授我统计课老师的专业背景不一样,其中有 2 个是数学专业出身，而且有的更多是注重背

后的数学公式推导，重在实现背后的讲解，软件实现涉及相对较少。同时，在上机课讲解软件实现的时候，只说了怎么勾选一些选项，怎么解读结果，而为什么不勾选另一些选项等问题则可能因为时间关系没有铺开讲。此外，加之缺乏文章写作基本经验，（那时不知道做计量模型前要先提出研究假设，然后去证实/证伪研究假设，更不知道要对数据做基本的处理，如处理极端异常值，对一些变量进行描述性统计分析，做散点图看趋势等等，只知道将自变量和因变量分别统计出来，然后就建模了）——这样做会出现很多问题，现在想想还有点后怕。这或许也是现在很多初学计量的人最容易犯的毛病，就这么一股脑地做下去了。

可能由于知识的不系统，故而经常做模型时前一秒做是一个结果，后一秒做又是另一个结果。到处寻求无果，当时我班上的同学不搞这个，没法交流。还好当时有机会下课后给老师请教，学到不少东西，这也迫使我去买专业的书籍，系统学习相关知识。

第三阶段：查缺补漏——计量似乎也不是想象中的那么难

由于本科学了一点 SPSS，加之这个软件交互式的操作页面，比较友好，故而首选这款计量软件。SAS 我在北京研究生院期间学了四门课，但比较零散，不系统，加之软件破解版都有 9G，太大了，所以慢慢的就把它弃了。后面因为科研需要另学了 STATA，但这属于后话了。

国内 SPSS 教材也算比较多的了，如何选成了一个问题。当时也没想那么多，

就在老师推荐的基础上买了本张文彤老师写的 SPSS 高级教程。通过教程我明白了很多细节的东西，例如，做回归分析时要勾选的框分别对应的是什么，也在这本书的指导下写了生平第一篇关于计量的文章。

在学习 SPSS 的同时，我在北京也经常去图书馆看书，也在网上买了些和我将来专业有可能相关的一些专著。慢慢地，我发现专著中出现的很多新名称，在我买的张文彤老师的 SPSS 高级教程涉及比较少，比如结构方程模型，异方差，WLS，2SLS，工具变量法，面板回归等。有的 SPSS 软件也操作不了或者不好操作，比如异方差检验，用 STATA 命令 BP 就可判定，而 SPSS 可能就要依照 BP 计算原理一步步去做了。那时也不知道该在哪里去找相关的资料，系统的学习一下。在网上搜索一些关键词，也只是零散的一点介绍，没有系统的。那时可能想着自己现在也不用这些方法，加之也比较忙，故而就搁置在那了。但还是乘机会买了万卷方法系列的几本好书，如《问卷统计分析实务--SPSS 操作与应用》、《量化研究与统计分析——SPSS(PASW)数据分析范例解析》、《结构方程模型——AMOS 的操作与应用》等。系统地学习了 SPSS 和 AMOS 软件，至此，SPSS 基本操作算是入门。但异方差性，WLS，GLS，FGLS 这些计量方法那时还是不知道怎么解决，因为买的这些书中基本都少有这些方法的介绍。有的只是一些基本的数据处理，简单的回归分析（OLS），因子分析，聚类分析等。

第四阶段：躲不过了，硬着头皮也要上

前面 3 个阶段，学了一些 SPSS 的知识，但似乎也只是建立在自己看书学习的基础上。现在是研究生了，不像本科时间那么充裕，有很多事要做，像导师的项目，自己写文章，日常的交际等。这样就导致很多知识，在看的时候理解了，但过段时间不怎么用就忘了。加之看书时，有时没有具体软件上手操作，所以后面要用的时候，都不知道要到哪里去找一个具体的小知识点了。在这时，有幸从一个师兄那里得到了张文彤老师 SPSS 内部培训视频（初级和高级），通过密集的视频学习，掌握了以前很多都没掌握牢固的技巧，进而使得 SPSS 运用更加熟练和得心应手。在此基础上买了张文彤老师《IBM SPSS 数据分析与挖掘实战案例精粹（配光盘）》这本书和相应的视频，进一步在实战中学习。至此，SPSS 基本操作对于我来说应该没什么问题了。但还是存在第 3 阶段那个问题——FGLS，2SLS 这些是什么我都不清楚。

经过一段时间看文献，有一天我找到一个 idea，去和导师商量后，觉得可行，打算写一篇好文章。可要写这篇文章需要用分位数回归。这个 SPSS 解决不了，STATA，EVIEWS，SAS，R 这些能解决。除了 SAS 外，我对其它几门软件一窍不通。在此情况下，我打算除了 SPSS 外，另学一门计量软件。于是百度比较，看哪个软件好学一些。很多人推荐 STATA，从此走上 STATA 学习之路。

由于 STATA 没有基础，怎么快速的学习它呢？有前面 SPSS 视频学习经验，果断去网上购买了连玉君老师 STATA 培训视频，至此努力学习。初级教程视频看

了一遍，没多大问题。到高级教程里面介绍 OLS 时也没多大问题，但到后面介绍 GLS，时间序列分析时，感觉听不懂了。其实也不是听不懂，是不懂这背后的机理，为什么要这么做。于是，搁置一段时间，直到那天——两本书的出现。

以前有个师姐打算考中科院农业政策研究中心张林秀老师博士，考试要考计量经济学，买了两本伍德里奇写的《计量经济学导论（第三版）》，后来因为英语差 1 分没上，直接出去工作了，这两本大块头书就遗留在办公室的柜子里。由于相对于经济学来说，我们是跨专业，加之平时比较忙，故而办公室的其他人对这两本书都是望而却步，将其束之高阁。有一天忙完事休息，我去翻柜子，突然发现这两本书，拿下来一看目录，竟然有 2SLS，异方差处理等介绍，那时真是喜出望外。这时也快要放寒假了，就想着寒假扛回家好好看看。

同时，由于也在学习 STATA，想着看看有没有 STATA 的工具书。通过人大经济论坛，找到了陈强老师写的《高级计量经济学及 STATA》以及另一个坛友说的郭志刚老师翻译的一本介绍国外 STATA 的经典书籍《应用 STATA 做统计分析》。下载了电子版，回家用手机慢慢看。

第五阶段：计量也不是想象中的那么难

可能是受长期中国教育模式的熏陶，我一直都比较急功近利。虽然寒假我将伍德里奇的书扛回家了，对于农村出身的我来说，转几次车拖一个箱子，箱子里转着这么两本大块头的书也是挺不容易，然而我最开始也不是去看它，而是看我

下载到手机的两本 STATA 工具书。看了两本工具书后，我想要解决的问题（分位数回归）至少在软件实现上我没什么问题了。但或许也是机缘巧合，也或许是农村过年无聊（很多我小时候的玩伴都出去打工了，没回来，就有一个人还在读博，走上了一条“不归路”），我拿起了这两本好不容易扛回家的大块头。

谁想才看没多少页，就有点看不下去了。虽然我大学学了《微积分》、《概率论与数理统计》以及《线性代数》，自认为也还学得不错。但时隔已经 6-7 年了，早就忘得差不多了。故而一看书中最开始推导 OLS 的高斯—马尔科夫假定就有点头大，但实在是因为无聊，硬着头皮把它看下来了。至此，我的学术生涯可能走向另一条不同的路了。

慢慢的，我越看越有劲。看着里面的很多案例，看着里面写的很多话真的是忍不住拍案叫绝，为什么写得这么经典，这么好。可能以前我也有一些底子，故而我看这两本论坛上大家公认的中级教程（初级是古扎拉蒂写的《计量经济学基础》，高级是格林的《计量经济分析》）也没费多大力就理解了。一些长久困扰我的疑惑，如自变量因变量取对数后模型怎么解释，2 期数据的面板回归怎么做等也慢慢的得到了解决。

经过这一系列过程，再回过头来看我这个外行的计量之路，似乎以前走的有点偏。以前是重实用，为模型而模型。现在更想做的是多探究为什么要用这个模型的机理，我想这可能才是一条长久之路。慢慢的现在也意识到了计量与统计的

可能不同，计量的博大精深。同时，也发现计量也不是想象中的那么难了。在国内大家觉得难可能真的是因为国内没有出像伍德里奇写的这么好的教材了（PS：以前在北京通过六维空间下载了西南财大庞皓老师讲的计量经济学视频，也买了他写的书，可能是限于当时的认知水平，觉得比较难——尤其是看到前面的 OLS 部分推导——故而没学多久就放弃了。现在回过头来看，此书也不失为国内少有的计量初级好教材。）

自此，自己也将怀着一颗虔诚的敬畏的心好好学习计量，用好计量。争取做一个地理学中的计量怪咖，我现在专业是人文地理学，搞农村聚落地理方面的研究。

啰啰嗦嗦又写了这么多，希望对想学计量但又苦于不知道怎么学的朋友有一点启示意义吧，也希望能与计量已入门，计量的大牛们多多交流学习。有机会一起合作文章也可以。

2.2

越努力 越幸运：自学计量的历程、心得、成败

原文地址：http://bbs.pinggu.org/thread-3880985-1-1.html

原文名称：一个外行的计量经济学学习之路二——经管答疑有感

论坛从答疑上线到后面的改版，专门确定经管答疑版块之“经管爱问”，到目前为止已有两个半月左右，有幸接受论坛的邀请，参与到论坛的答疑中来。答疑两个半月，前前后后答了几百个问题，尝试着总结一点提问者共性的东西，结合这段时间我这个外行计量学习的心路历程变化，写了这个帖子，与大家讨论学习一下。不当之处欢迎批评指正。

我真正意义上入驻论坛的时间不长。注册时间是 2013 年，真正意义上活跃在论坛是从去年年前左右开始。当时入驻的目的可能和现在大多数坛友相同：就是遇到计量/统计上的难题了，迫切想论坛的大神们帮忙解答下，要是能手把手地帮忙指导，或是远程辅导，甚至帮忙做一下数据分析更好。然而正如我在“一个

外行的计量经济学学习之路”中说的一样，入驻一段时间，发现自己想解决的问题没得到有效解决，反而在这个热心的过程中顺手帮许多坛友解决了问题，虽然由于自己的水平有限，很多问题都是初级的问题，但还是收获颇丰。所以慢慢的我知道：求人不如求己，很多东西是需要自己去琢磨的。要想取得成果，非下苦工不可。

后来机缘巧合之下，发现了伍德里奇的《计量经济学导论（第三版）》，整个寒假的时间沉浸在这两本书里面，前前后后看了三遍（第一遍是硬着头皮来的），慢慢的开始上路了。后来又有幸在论坛“蓝色”版主回一个坛友的帖子中发现了谢宇老师的《回归分析》一书，通过此书发现谢宇老师的另外两本好书：《社会学方法与定量研究》和《分类数据分析的统计方法》；再到后面课题组有个师姐写论文要用到 HLM 软件处理分层线性模型，接触到张雷老师的《多层线性模型应用》，郭志刚老师翻译的《分层线性模型:应用与数据分析方法》等书。我算是如获至宝，第一时间利用各种渠道买到纸质版（个人习惯，喜欢看纸质版的书，看书时可以做一些笔记在上面），网上没有纸质版卖的直接找电子版，然后到打印店打印出来看。拿到书后，集中时间和精力，先将书快速而系统的过一遍，觉得好的地方再进一步细看。同时，看的过程中不忘写点文章练手。其中，用到了分层线性模型、分位数回归等以前没有用过的方法，近期打算用面板回归处理一篇文章。另外还有一本诚意推荐，就是上面忘了提到的陈强老师的《高级计量经济学及 STATA 应

用》，这本书真心写得好，我前后看了 4 遍，但由于这本书语句比较精炼，没有一点基础去看可能会很艰难。即使有了一定的基础，有很多地方可能还是需要时间才能很好的理解。

我时常在想，我一个计量/统计外行，通过自学出道慢慢的入门了，要是换一个人来，如果走我相似的路（如看这些书），是否会一样呢？

前几天，和我论坛的导师（现实生活中的师弟）“我的素质低”以及和论坛编辑（现实生活中的师姐）聊天。我说我是个计量外行，自学出道。大概有个重软件实现轻背后理论——恶补理论——理论与软件实现并重的这么一个心路历程。这里之所以把理论放在软件实现前面，是因为现阶段的我认为计量理论是根本，是“道”；软件实现是工具，是“术”。唯有两者并重，可能才能走得长远。

不过话又说回来，计量理论掌握好了，一通百通。软件的实现慢慢就会变为其次了。就像我自身学习 STATA 半年多时间，发现 STATA 的计量模型实现过程和结果解读，很多可以迁移到 Eviews 上去，所以现在如果让我去学 Eviews，我想应该会很快上手。同时，和师姐聊到现在博士阶段，作为导师身边的主力，比较忙，不会再像去年寒假时有那么多时间集中的再来学习计量了，只能零散地学习。师姐提到了一个观点，我个人比较认同，可能适用于现在大多数来论坛计量版块提问的坛友们。

师姐的观点大致如下：“计量的学习是个系统长期的过程，如果只是用零散的

时间去学，可能掌握的知识点也是零散的，不系统的。可能是不利于长期发展的”。

我个人的经历大概也是这样。就像我上段话中描述的一样，我看伍德里奇的《计量经济学导论》是利用寒假的时间集中看的，第一遍是真心辛苦，尤其是这么多年没碰《线性代数》、《概率论与数理统计》和《微积分》，所以看到高斯马尔科夫假定，看到 OLS 系数推导等地方时，真心觉得有点看不下去的感觉，还好实在闲得无聊坚持了下来。

我看陈强老师《高级计量经济学及 STATA 应用》，看谢宇老师的几本书，看分层线性模型方面的书都是集中时间看的，长的 10 天左右。短的可能就 2-3 天第一遍就完了。前后一鼓作气地看有利于集中精力，且前后相对连贯。同时，持续一段时间的集中看书学习，容易找到一种看书的状态。这种状态找到了，可以帮助持续地沉下心来看下去。

现在忙了，好不容易找点空闲的时间看看书，却不一定进得了状态；好不容易进了状态，结果又通知有事要做，导致好不容易找到的状态瞬间没了。等有时间了却不一定有看书的那个状态了(即真正静下心来潜心地系统学习研究)。所以，我个人的经历能给后来人的启示可能就是：找好适合自己的资料，找到看书学习的一个状态，集中精力的来系统学习。何为适合？在我看来，就是拿着一本公认的经典书籍，看看自己现在的基础能否看得懂，在多大程度上能理解。就像看计量的书籍一样，零计量学习经历，但有一定的数理基础的可能看古扎拉蒂的《计

量经济学基础》较为合适，有一定的计量基础的可能看伍德里奇的《计量经济学导论》较为合适，基础深厚的可能看格林的《计量经济分析》较为合适。不同的基础对应不同的书籍，我的感受是，找准找对书籍对后面的学习很重要。

可能没基础的人最初看书会看着头大，看着吃力且看不进去。这时，有一定的感性认识可能会好一些。如果是这样的话，经济条件好的建议报论坛的计量/统计软件培训班。张文彤老师的 SPSS 培训和连玉君老师的 STATA 培训视频我都看过，系统地学习过，真心比较好。参加培训班的好处是即使培训后对计量/统计理论还是一知半解，但会多一层感性认识。举个例子，SPSS 统计教程书上的前几章内容，例如软件功能介绍，数据导入等基础知识。对于没接触过 SPSS 软件，无计量统计学习基础但又想自学的人，可能会看着就头大，搞半天搞不出个所以然来。但如果参加培训，看几天都没明白的东西可能经过 1 个多小时的培训课程就明白了。如果是边参加培训边实践操作，那认识自然会更加深刻。有了这层基础，再回过头来看经典教材，可能就相对容易得多了。我想，我之所以看一些计量/统计教程速度会那么快，与看过以上两位老师的视频教程是有莫大关联的。

经济条件差的建议在论坛找找以上我说的书的电子版（"一个外行的计量经济学学习之路"，这个帖子中也有很多书），然后找准适合自己基础的书，看电子版或买一本来系统学习。同时，现在网上也有很多分享计量/统计学习的视频，也可找一些免费的来看看，学习。

然而，可能需要注意的是，不是买了视频或者参加了培训就会变得很厉害，就算计量统计入门了。毕竟培训老师讲得再细再好，那也是老师的知识。要系统地转化为自己的东西才好。师傅领进门，修行靠个人的。这里其实又涉及到一个集中学习的过程了。我个人的经验是：看视频尤其是第一遍视频，即使看不懂也尽量耐着性子硬着头皮看下来，不懂的地方看第二遍第三遍；看的同时，跟着老师的思路走，边听边自己操作。同时，视频系统学习后回过头来集中学习软件实现背后的理论知识。

我个人的亲身经历是在看连玉君老师 STATA 视频高级部分时，那个时候计量理论知识比较缺乏，而连老师又讲得很精炼，所以当时的感觉是听着很吃力，有点像是听懂又像是没听懂的感觉。说听懂了，是老师讲的通俗易懂，貌似是那么回事；说不懂呢，是因为老师讲的都是高度浓缩的知识点，如果不去看书，没法将这些知识点很好地串联在一起，所以就感觉比较吃力。当时遇到的具体情况是，看初级部分视频没问题，看高级部分视频的 OLS 回归部分没问题，但一到时间序列部分，面板分析部分，我就有点头大了，不得已只好放弃。后来系统地学了伍德里奇《计量经济学导论》和陈强老师《高级计量经济学及 STATA 应用》，慢慢的对计量理论的理解更深刻一些，再回过头来看视频就没那么吃力了。同时，高度概括的视频讲解对相对繁琐的书籍理论介绍进行了提炼，使得自身理解更上了一层楼。

据我了解，即使伍德里奇的《计量经济学导论》现在网上也有培训视频卖，不可谓不是广大计量求知爱好者的福音。总结而言，软件培训更多的是帮助我们增加对软件及软件操作的认识，当然对计量理论的学习，也有一定的感性认识上的帮助，然而具体深入的计量理论学习可能还得靠自己私底下静下心来扎扎实实地看，不断地操作实践，进而不断的进步。

后来，我又在想，即使计量理论掌握得比较扎实了，基本的软件操作也会了，可以用软件快速的实现各种所谓的计量模型，那么就一定能在实战中所向披靡么？这也未必。可能是因为书籍写作章节安排以及软件培训分块相对系统的原因，即使是经典的计量统计书籍，很好的软件培训，知识点也只是相对集中的，对于建立一个解决实际问题的模型而言（如写毕业论文，发文章），可能就是零散的了。

举个具体的例子：我们要构建一个常见的横截面数据的 OLS 回归模型（即多元线性回归模型）经典的计量统计书籍及软件培训可能会相对集中的分版块分章节告诉我们要构建这么一个模型，需要的数据类型是什么样的（如因变量是连续性变量，自变量如果是类别变量需要处理为哑变量），有一些什么样的基本假定（如高斯马尔科夫假定），结果怎么解读等。即构建这么一个模型的所有细节的东西可能分版块（章节）讲解了，但没有集中起来。然而我们在实际写文章/毕业论文的过程中，是需要一个系统的思维过程，是需要把这些零散的知识点整合起来。

比如现在如果要我构建一个横截面数据的 OLS 回归模型，我可能会考虑以下

的东西（这里只做简单的介绍，详细一点的介绍请参见我的另外两个帖子〖"万万"个论坛币〗计量实证的“摸底”考试和 SSCI/SCI 发表的一点心路历程）。

一是构建模型的指标怎么来。在稍微好点的国内外期刊中，模型指标的选取是基于理论/文献/有理有据的常识来的，并会在构建模型前做严格的研究假设。

二是构建模型前数据的预处理。如做描述性统计分析看看数据是否存在极端异常值；数据是否基本符合正态分布，如果不符合怎么处理（如对于含 0 较少的右偏态分布数据取对数；对于含 0 较多的右偏态分布数据将连续性变量处理为类别变量）；是否将类别变量处理为哑变量；对数据变量间是否存在严重多重共线性进行诊断等（曾经尝试写了这么一个帖子，是关于极端异常值诊断的。描述性统计分析在实证研究中的作用及具体软件实现——以 SPSS 为例）。

三是对数据进行建模，然后对模型进行合理的解释。建模时考虑结果是否稳健，是否存在内生性变量等。

之所以会举这个例子是因为我在答疑过程中发现：很多坛友都是拿到数据直接就到第三步的建模过程了，淡化了第一步指标的选取过程和第二步数据的预处理过程。这是非常可怕的一种做法。然后我又结合来提问的坛友的可能特征想了想，至于为什么这样的原因就了然了。经过我个人的观察，来提问的坛友可能大多具有以下特征：

一是大多是本科生或者硕士生，未学过计量统计或懂一点“皮毛”的居多。

真正计量功底深厚的人大多潜水了，自己有问题基本自己就解决了。这部分人即使有问题，也知道通过何种渠道去找资料，能快速的找到相应的资料解决问题；还有一种是介于以上二者之间的，恰如我。懂一些计量，能解决一些基础问题，且正好又比较热心，所以被论坛 HR 挖掘过来当版主，帮助大家答疑解惑了。

二是对一些计量统计理论的理解都不深刻，问的问题确是相当高深复杂的模型。这部分人层次跳跃很大。举个我答疑中的例子，有个坛友不知道 z 统计量是什么东西，问的问题却是面板 Logit 模型，着实吓了我一跳。

三是构建的模型多是模仿着文献中的来，自己找同样的指标构建同样的模型却得不到一样的结果。我在第二点举的例子中的坛友可能就在此列。

四是有一腔热血想快速地学好计量。这类坛友经常问的问题是如何学好计量/统计/某门软件操作，要多少时间才能学好之类的问题。

我时常在想，对于来论坛提问的大部分坛友，他/她们中可能只有很少一部分人会继续深造，或者出去工作后会用计量/统计相对较少。更多的坛友可能只是面临写文章/毕业论文的压力，而迫切的需要用到计量/统计软件。他/她们中的大多数可能把论文写完了就出去工作了，不会再接触这些所谓的计量模型了。

同时，很多坛友基础知识都没掌握好，一来就问很高深的计量模型，这使得我们这些所谓的答疑“专家”或者热心答疑的坛友很尴尬。他/她们的模型构建大多模仿着文献来，软件操作仿着书来。由于缺乏一定的基础，我们即使回答了他/

她们的问题他/她们也不一定看得懂，因为很多时候计量模型的建构是个系统的过程，难免会涉及到一些专有名词，即使是一些很基础的专有名词。我记得我的答疑中有用复杂的计量模型却不知道多重共线性是怎么一回事的。更有甚者，有的坛友想当然的，“天马行空”式去构建计量模型，指标的选取脱离了理论/文献/基本常识。这使得我们答也不好，不答也不好，很尴尬。对我个人而言，我是本着一颗热情的心，只要我会的还是尽可能的帮助，并会在回答中提到要注意指标的选取和数据的预处理。

我系统地梳理了下我的几百个答疑，发现只要是涉及到回归方面的问题大多是没有进行我上面举的例子中提到的第二步。故而常见的问题是模型的回归系数方向与参考的文献相反、或与自己预计的方向相反、或与自己做散点图发现变量间的方向相反。这些问题多是极端异常值/变量间的严重多重共线性引起的。很多坛友经我一说后，马上去诊断一下，很多就得到自己想要的结果了。

我一直在不断地反思自己这一路的计量自学过程，以前写文章到处碰壁的过程。深感计量理论的学习，软件实现，文章的写作，投稿修稿到最终稿件见刊的这一系列过程是一个系统的过程。任何一个环节弱了可能都会使得文章经历一些坎坷后才能见刊。在这反思的背后又对前面提到的部分坛友的情况感到担忧：文献没读几篇，缺乏一定的鉴别能力就匆忙地找篇范文模仿着选取指标了（可能参考的文献模型建构的都是有问题的）；指标的选取没有基于一定的理论/文献/

客观有力的常识，而是“天马行空”想当然的就选取了，且在选取时贪多求全，忽略了模型的简洁性；指标选取后没做严格的理论假设，没对数据做基本的预处理就模仿着文献和软件实现书籍仓促的建模了；有的“更可恨”的是明明是建模过程中出现了问题却为了迎合主观常识，人为的篡改了模型数据（这样的结果可能行家一眼就看出来了）。

担忧归担忧，但我也知道所有事物的转变总有一个过程，我自己前期写的中文文章也存在这样或那样的不足。对于前面提到的只是暂时需要帮助，事后工作后与计量接触不多的这部分坛友我也表示理解。正因为理解，所以看到有坛友提问，只要我能答的，也会尽可能的在经管爱问中给予解答。写此贴的目的一是在于再次回顾自己计量学习的一点心路历程，希望给想学好计量却又苦于不知道如何学的坛友一点启示；二是简单的归纳总结了下答疑两个半月以来的一点心得和一点担忧，希望看到这个帖子的坛友（尤其是基础不那么好的坛友）在以后构建模型时多想想我举的例子的第一步和第二步（有时间我再开一贴对这一系列过程做个系统的论述）。前期准备越充分，后期才不会有那么多不确定，那么多烦恼。

想说的话还有很多，但一看这篇幅，想想还是算了。留待以后吧。需要说明的是，此贴更多的只是在心路历程上和一些“假大空”的实现上进行了一些梳理，具体的计量理论和软件实现过程还是得靠自己私底下辛勤的付出才行。

越努力，越幸运。与君共勉。

2.3

SPSS 数据分析心得小结与分享

原文地址：http://bbs.pinggu.org/thread-3569928-1-1.html

原文标题：SPSS 数据分析心得小结_数据分析心得分享

学习数据分析，会经常用到 SPSS 分析工具，这可真不是一般的功夫，需要很认真、很细心才能运用好它。下面我来和大家分享一下关于 SPSS 数据分析的心得小结，希望大家从这些亲身实践经历中有所收获。

心得 1：拿到一份数据，或者在看到国内外某个学者的文章后有想法，并且自己手里的数据刚好符合这个想法可以做时，在整理好数据后不要急于建模。一定要对数据做缺失值处理、异常值处理。在数据预处理的基础上再进一步建模，否则可能得到错误的结果。

心得 2：承接心得 1，数据预处理怎么做。一是缺失值的处理。我个人有几个看法：一是数据样本量足够大，在删除缺失值样本的情况下不影响估计总体情

况，可考虑删除缺失值；二是数据样本量本身不大的情况下，可从以下两点考虑：首先是采用缺失值替换，SPSS 中具体操作为“转换”菜单下的“替换缺失值”功能，里面有 5 种替换的方法。若数据样本量不大，同质性比较强，可考虑总体均值替换方法，如数据来自不同的总体（如我做农户调研不同村的数据），可考虑以一个小总体的均值作为替换（如我以一个村的均值替换缺失值）。其次是根据原始问卷结合客观实际，自行推断估计一个缺失值的样本值，或者以一个类似家庭的值补充缺失值。

心得 3：关于数据预处理第二点异常值的处理。我大概学了两门统计软件 SPSS 和 STATA。SPSS 用的时间久些，熟悉一下，STATA 最近才学，不是太熟。关于这两部分我尝试结合着来说。关于异常值的处理可分为两点，一是怎么判定一个值是异常值，二是怎么去处理。

判定异常值的方法我个人认为常用的有两点：一是描述性统计分析，看均值、标准差和最大最小值。一般情况下，若标准差远远大于均值，可粗略判定数据存在异常值。二是通过做指标的箱图判定，箱图上加“*”的个案即为异常个案。发现了异常值，接下来说怎么处理的问题。大概有三种方法：一是正偏态分布数据取对数处理。我做农户微观实证研究，很多时候得到的数据（如收入）都有很大的异常值，数据呈正偏态分布，这种我一般是取对数处理数据。若原始数据中还有 0，取对数 ln(0)没意义，我就取 ln(x+1)处理；二是样本量足够大删除异常值

样本；三是从 STATA 里学到的，对数据做截尾或者缩尾处理。这里的截尾处理其实就是同第二个方法，在样本量足够大的情况下删除首尾 1%-5%的样本。缩尾指的是人为改变异常值大小。如有一组数据，均值为 50，存在几个异常值，都是 500 多（我这么说有点夸张，大概是这个意思），缩尾处理就是将这几个 500 多的数据人为改为均值+3 标准差左右数据大小，如改为 100。总结而言，我个人认为做数据变换的方式比较好 数据变换后再做图或描述性统计看数据分布情况，再剔除个别极端异常值。

心得 4：如何做好回归分析。经过多次实战，以及看了 N 多视频，上了 N 多课，看了 N 多专业的书。我个人总结做回归的步骤如下：一是承接心得 1-3，对数据进行预处理，替换缺失值和处理异常值；二是将单个自变量分别与因变量做散点图和做回归，判定其趋势，并做好记录（尤其是系数正负号，要特别记录）；三是自变量和因变量一起做相关系数，看各个变量相关关系强弱，为下一步检验多重共线性做准备；四是自变量严重多重共线性诊断。若变量存在严重多重共线性，可采用主成分回归，即先将存在严重多重共线性的变量做主成分分析合并为 1 个变量，然后再将合并成的新变量和其余自变量一起纳入模型做回归；五是做残差图，看残差图分布是否均匀，一般在+-3 个单位之间均匀分布就比较好；六是报告相应结果。

心得 5：看到论坛上有网友问为什么他（她）们的老师不建议采用后向步进

法处理变量严重多重共线性。记得张文彤老师说过他有个同学做过一个研究，即采用后向步进法剔除变量的方式去做回归，得到的结果犯错的几率比较大。张老师也不建议用这个方法处理严重多重共线性。处理严重多重共线性比较好的方法是做主成分回归。

心得 6：有个朋友问我在报道回归结果时用未标准化的回归系数好，还是用标准化后的回归系数好。我个人觉得这个问题仁者见仁智者见智，要看想表达什么。具体而言，如果想表达在其它条件不变的情况下，自变量 X 每变化 1 个单位，因变量变化多少个单位，这种情况用未标准化回归系数就好；如果想比较各个自变量对因变量影响的相对大小，即判断相对而言，哪个变量对因变量影响更大。这时需要消除量纲的影响，看标准化后的回归系数。

心得 7：这是投稿一篇 SSCI 外审专家提出的意见。我做的是无序多分类 logistic 回归模型。因变量分了 5 类，有一类个数比较多，达到 300 多，有 1-2 类个案比较少，只有 30 左右。专家提到了要做稳健性检验。这个用 STATA 软件编程加一个 robust 即可解决问题。不知道在 SPSS 里面怎么做。欢迎知道的朋友一起讨论下。我个人认为这是一个好问题的。不做稳健性检验模型可能受一些极端值的影响，结果不稳定。可能本来显著的变量剔除 1-2 个样本后就变得不显著了。所以做回归分析稳健性检验也比较重要。

最后，如果有可能，我希望在后面的心得中附上实际操作的步奏图和解释。

看看大家是否需要。不然，可能说的一些东西需要的人能看明白，但还是不清楚怎么做。也希望和大家进一起交流学习。

2.4

学习地图：如何测试自己的计量实证水平

原文地址：http://bbs.pinggu.org/thread-3641714-1-1.html

http://bbs.pinggu.org/thread-3641724-1-1.html

原文标题：测测您的计量实战水平——经管代码库诚邀各位英豪

引子：我们学习计量的目的是什么，相信大多数人大经济论坛的坛友都是为了解决毕业论文，发文章的问题。那么，您真的懂计量么？如果您认为您懂，请您看看下面我根据个人实证经历整理的一些问题，看能否有效的解决呢？本文是一个简单的测试和练兵，希望以顺序渐进的提问方式，梳理提高实证水平之旅。

1、 拿到一份数据资料，您能快速地判定这是什么类型的资料么？请举例说明横截面数据、时间序列数据、面板数据等的区别；同理，拿到一份数据资料，您能快速地判定变量类型么？请举例说明类别变量、有序变量、定比变量的差别。

2、 数据类型和变量类型明确后，接下来是方法的选择了。请举例说明不同数据类型、不同因变量类型所对应的方法吧，例如，只写横截面数据，那么因变

量对应于连续变量、类别变量分别用哪些方法呢？请尽可能的详尽哦（此处仅列举方法即可哦，不必将方法延伸了来讲）。

3、 在做计量分析前，一般要做描述性统计分析（报告均值，标准差，最大值和最小值），做这个分析的目的是什么呢，欢迎大家深入探讨，敞开想吧。同时，如果我有 3 个变量，分别为受教育年限，性别和婚姻，您能快速地利用软件做出不同年龄、性别下个体受教育年限的频数分布表么？

4、 一份数据资料（尤其是微观实证资料），一些指标经常会出现极端异常值，请问您注意到它了吗？有哪些手段和方法可以帮助我们去发现极端异常值呢？请举例说明；如果发现有极端异常值，而我们没处理，那会对模型结果产生什么样的影响呢？请举例说明；如果不小心和它偶遇了，我们怎么解决它呢？如果变量中不小心有大量的 0，且成正偏态分布，这种情况我们又怎么解决呢？这个问题理清楚了，对实证分析很有帮助哦，快来解决吧。

5、 高斯—马尔科夫假定一般是做计量的基础，那么它有哪些假定呢？请将其假定写出来，并简要说明假定组合的用处吧。

6、 变量的独立同分布，同方差，且服从正态分布，是很多模型的假定基础，那么您知道有哪些方法可以判定一个变量是否服从正态分布呢？请举例说明吧。

7、 高级点的问题来了：您是否听说过异方差呢？如果模型一不小心出现了异方差怎么办，会对模型结果产生什么样的影响呢？我们有哪些手段可以去判定

一个模型是否存在异方差呢？如果发现模型确实存在异方差了，我们有哪些办法可以去解决它呢？请举例说明，欢迎继续摸索。

8、 更高级的问题来了：您是否听说过变量的内生性呢？如果模型一不小心出现了内生变量怎么办，会对模型结果产生什么样的影响呢？我们有哪些手段可以去判定一个模型是否存在内生变量呢？如果发现模型确实存在变量内生的情况，我们有哪些办法可以去解决它呢？请举例说明。

9、 承接第 8 个问题，如果您找到了工具变量去解决变量的内生性问题，那么您怎样判断您找的工具变量是合理的呢（不要告诉我文献是这样处理的，哈哈）？我们有哪些方法可以去判定工具变量选取的合理不合理呢？欢迎举例说明。欢迎来挑战。

10、 计量模型中，回归分析是重中之重。那么我们做了一个回归分析后，要报告哪些结果呢？请简要做一个案例，将您认为需要报告的东西列举出来，并做简单的解释吧。同时，请考虑下 R^2 是不是在所有模型中都很重要（如多元线性回归的 R^2 和 Logistic 回归的伪 R^2,它们是一样的么？ ）。同时，请再考虑有没有截距项对模型有什么样的影响呢？都是些有挑战的题目哦，请举例说明。

11、您做完回归分析后，怎么让别人信服您的模型是合理的，是最优的模型呢？如果有人告诉您您这个模型可能遗漏了某几个比较重要的变量（对因变量影响显著），模型设定的有问题，您怎么回答这个问题呢？请举例说明遗漏变量对模

型结果的影响（这个也可以给高奖励哦）。同时，我们在文献中经常看到有些变量取了对数，有些变量加了高次项（如 2 次），您认为作者是凭什么对数据进行这样的处理呢？请举例说明。您的毕业论文或者发表的文章有考虑过这个问题么？这是一个比较有挑战的问题哦，快来挑战吧。

12、 再来谈点细节的问题，回归分析中，您知道自变量对因变量的解释是怎么样的吗？如因变量是连续变量，自变量是连续变量，两者做回归；如果因变量取了对数，自变量不变，要怎么解释呢？如果因变量不变，自变量取了对数，又要怎么解释呢？如果因变量和自变量都取了对数，模型又要怎么解释呢？请举例说明。有高奖励哦。同时，如果我们做的是 Logistic/probit 回归，那么回归系数怎么解释呢？优势比（OR）又要怎么解释呢？边际效应（均值的边际效应和在任意个体值的边际效应）怎么解释呢？此外，模型中自变量是连续变量和类别变量，解释有差别么？欢迎举例来说明哦。有高奖励。

13、漏问了一个很重要的问题。您知道多重共线性么？万一模型不小心出现了严重多重共线性怎么办？出现严重多重共线性会对结果产生什么样的影响那？请举例说明。我们怎么去判定一个模型是否存在严重多重共线性呢？如果真存在了，有哪些方法或手段去解决它呢？这些解决的手段有优劣么？欢迎举例来说明。

14、常有坛友问因子分析和主成分分析的问题。那您知道这二者有什么区别么？我们用这两个方法可以解决什么问题呢？如果我有 15 个指标，我想要 3 个

主成分，而利用因子分析/主成分分析做我只得到 2 个主成分，怎么解决呢？您考虑过如果一个主成分特征值为 1.1，一个主成分特征值为 0.9，我们怎么处理这个问题么？是像一般书上说的特征值 1 一下的我们直接舍弃，还是根据具体情况两个都保留呢，您怎么看这个问题。欢迎举例说明。有高奖励哦。此外，主成分回归您听说过么？它能很好地解决严重多重共线性的问题哦，欢迎举实际案例来具体说明吧。有高奖励哦。

15、 物以类聚，人以群分相信大家都听说过。那聚类分析您会做么？对变量聚类和对案例聚类有什么区别呢？如果我们想聚类的变量为正偏态分布的变量，我们可以做横向标准化的中位数聚类，这个您会做么？欢迎举例说明。有高奖励哦。

16、还有些我个人认为平时我们不常接触到（如 SPSS 国内的书籍一般没涉及），但应用却越来越热的模型，如分位数回归，泊松回归，负二项回归，断尾回归，归并回归，tobit 模型等，您听说过么？它们是要在什么情景下使用呢？请具体举例说明，科普推广一下哦。这个有高奖励。

分界点：以上问题您可只针对横截面数据来看，下面我们来看看面板数据的问题哦，欢迎大神来解答啊。需要说明的是，由于我对时间序列这块接触的少，不是很了解，故而这块的问题相对较少。欢迎懂行的大神补充。

1、面板有平衡面板和非平衡面板，那您如果拿到的是一份非平衡面板数据，

您怎么将其处理为平衡面板数据呢？请举例说明。有学者说将非平衡面板处理为平衡面板，会损失很多信息，建议谨慎处理，您怎么看这个问题呢。欢迎尽情讨论。

2、说到面板回归，您知道解决一般面板线性回归的几种方法么？混合 OLS 回归，固定效应模型和随机效应模型这三个方法有什么区别么？我们怎么判定该用哪种方法呢？欢迎举例来说明。

3、按照 N（个案）和 T（时间）的长段，面板可分为长面板（T>>N）和短面板（T<<N），当然还有 N 和 T 相近的面板。这几种类型数据对应的面板计量各有什么特点呢？分别用什么样的命令或检验去实现一些特殊的要求（如异方差、截面相关、自相关和内生性呢）呢？欢迎举例说明。

4、同横截面数据一样，面板数据也要面对多重共线性、内生性、序列相关、横截面相关和异方差等问题，我们怎么去判定面板数据存在这些问题呢？如果发现了存在这些问题我们怎么解决呢？请举例说明，这个问题会有很高的奖励哦（这些问题可具体拆分为几个小问题来分开解答）。

5、同横截面数据一样，面板数据也会面临方法的选择。如因变量为连续变量、类别变量、有序变量对应的方法不一样，具体有哪些方法呢？请举例说明。如果能说明下利用计量软件如何实现这些问题更好。要很高很高的奖励哦。

分界点：考虑到来论坛求助的很多坛友都是本科生和硕士生，可能接触到的

计量的东西相对更集中在横截面数据和简单的面板数据。故而根据我个人经验整理的问题大概就是以上这些，欢迎大神继续补充并解答。此外，我个人一直认为计量软件实现是一个层次，掌握背后的原理又是另一个层次。故而，本版块特别欢迎坛友分享个人计量知识或者软件实现学习心得（不管你是初学者还是有一定基础的人还是大牛，鼓励原创，高奖励）。借用我导师一句话："学问是交流出来的，不要自己一个人在那闭门造车，多出去见见世面，多与人交流"。因为我们年轻，所以有无限的可能。奔跑吧，少年们。

结尾处：推荐几本我这个外行慢慢入门的关于计量学习的书吧。

SPSS 方面的：张文彤老师的《SPSS 统计分析教程-初级篇》、《SPSS 统计分析教程-高级篇》、《IBM SPSS 数据分析与挖掘实战案例精粹》；吴明隆老师的《问卷统计分析实务——SPSS 操作与应用》、《结构方程模型——AMOS 的操作与应用》；邱皓政老师的《量化研究与统计分析——SPSS(PASW)数据分析范例解析》；郭志刚老师的《社会统计分析方法——SPSS 软件应用》、《Logistic 回归模型——方法与应用》。这些教材基本我都买有纸质版，都学习过，各有各的优点，仅供需要的朋友参考。

STATA 方面的：陈强老师的《高级计量经济学及 STATA 应用》；汉密尔顿著，郭志刚等译《应用 STATA 做统计分析》；陈传波老师的《STATA 十八讲》。如果英语比较好，可以看国外相关专著，或者直接看 STATA 帮助手册。

需要说明的是，由于我个人只学习了这两门统计分析软件（Sas、Matlab 这些虽有所接触，但没买过书）故而对学习其它计量软件的朋友可能帮助相对较少。请见谅。正如我上面所说，学习计量掌握原理比软件实现更重要，所以推荐几本计量经济学公认的好书:古扎拉蒂《计量经济学基础（第五版）》——计量经济学初级教程；伍德里奇《计量经济学导论（第四版）》——计量经济学中级教程；格林《计量经济分析（第六版）》——计量经济学高级教程。将教程和软件实现工具书一起学习，会取到事半功倍的效果。我就是受益者之一。如果您真的想学好计量，要好好搞学术，不要在意自己的数学功底到底怎样，赶快买书或者下载电子书学习吧，会有意想不到的效果。

2.5

ssci/sci 发表的一点心路历程

原文地址：http://bbs.pinggu.org/thread-3695264-1-1.html

写在前面的话：本人是学农林经济出身的，主要做的是农户方面的微观实证研究，故而以下谈的个人文章发表经验可能不适于其它一些专业（毕竟各个专业有其自身的特点），但有很多东西应该有共性。写这篇帖子的初衷是以前在论坛发表过一篇个人中文投稿心得的帖子，从评论来看，对很多坛友有启发意义。所以在此将自己一路英文文章发表的历程也和大家分享一下，希望对初期投稿的坛友有所启示。不当之处请大家批评指正。

我是中科院一个研究所的普通学生（硕博连续了，现在博一）。在写本文之前，我已经有了几篇中文文章的发表经验了。由于在中科院博士毕业需要一篇SCI/SSCI/EI 源刊+1 篇北大中文核心，所以在导师的几番催促下，想着早点有篇英文文章好早点达到毕业条件，不用到博三快毕业的时候为文章发愁（PS：发一

篇 SCI/SSCI/EI 源刊+1 篇北大中文核心这个政策在我这个研究所才实行 1 年左右，可能当时所里通知的比较晚，所以上届的博三有 1/3 左右的师兄师姐延期了）。于是，我从去年 8 月份去三峡库区调研获得数据后就开始看文献着手准备写文章，前后写了 3 篇文章，下面我就把这 3 篇文章经历穿插起来，做一个经验总结。

由于是生平第一次写英文文章，加之对自身英语水平也不是很自信（毕竟是屌丝出身，不像北上广等大都市培养的学生很多英语都很好），所以开始写都是半中半英的，且模仿痕迹很重。写完后再用灵格斯软件慢慢翻译。为什么说模仿痕迹很重呢？现在回头来看，中国人的一些思维方式和表达习惯和外国人还是存在一些差别的。即使是地理口（我本身专业是人文地理学，现在主要研究方向是农村聚落方面的）的一些好的中文期刊（如《地理学报》、《地理研究》、《地理科学》等），其微观实证方面文章的质量和国外的一些 SSCI 还是存在一些差距的（如中文期刊的研究假设写得很简单，有的文章不会写 Theoretical Framework，研究结论不会与国内外其它类似研究做对比，不会深入分析与研究假设相反原因等。当然，这些差别都是我文章被拒多次后，终于有个很负责的期刊给我审稿意见我才慢慢发现的，后文会慢慢阐述）。所以，最开始我写英文文章的时候，心里面是没有底的。因为中文文章都还没写几篇，突然就要去写英文了，还想着 SCI/SSCI。然而无知者无畏。我手里面的微观实证数据是关于农户生计方面的，所以我用 web of science 数据库对农户生计方面的关键词进行检索，然后去一些顶级期刊

上找相关文献（经典的+最新的）。我这个方向，比较好的期刊如 Agriculture Systems, Food Policy, Journal of Rural Studies,然后把几篇用得着的经典文献打印出来，仔细品读，并作笔记。同时，为了搞清楚一些概念/研究方法，在当当上去搜索农户生计方面的书籍，然后买了几本比较好的中文专著，并仔细耐心地看了几遍，并做好相应的笔记（这对后面研究假设的提出很有作用，后面会细说）。在此基础上，把几篇经典文献中个人认为比较好的设定，如文章结构安排抽离出来，作为我自己文章的设定。框架有了，下面就是写了。

我最开始还是按照以前写中文文章的思维去写的。结果引言部分写的很宏观很大，这貌似是很多国人写文章的“通病”。然后文献综述部分把国内外很多类似的研究成果堆砌在那里，并没有做进一步深入“综述”，相当于只做了肤浅的汇总，没有进一步讨论，指出研究与其它研究的差别，指出研究的贡献，所以后面被审稿人提了，当然这是后话。然后是研究方法部分，这部分用了一个无序多分类 Logistic 回归，因为有很多文献介绍，当时没多大问题。然后我就直接进入数据描述性统计分析和结果分析了（没有做研究假设）。最后，自然是结果分析（没有和其它同类研究做对比），结论和政策建议（中文文章惯用的套路）。

从文章整理数据到写作初稿完成，3 周左右时间。当时写完也比较有成就感，没有进一步润色打磨一下文字，就直接“显摆”的发给导师了。因为导师以前也很少搞这方面研究（导师以前是搞自然地理出身的，后来转到搞人文这块，其实

他的专长在搞生态，我算是他学生中的一朵“奇葩”)，且对计量经济学这方面东西也懂得不多，故而就从他自身多年对研究区的感性认识出发，给我提出了一些很好的修改建议。然后就让我翻译了投出去。那时，我想，原来写 SCI 也不是那么难嘛（看我这么快就搞定一篇了）,后来我才发现那时的我是多么的 too young too simple。在信心极度膨胀之下，我在前面一个博士师姐的指导下，开始了我漫长的 SCI/SSCI 投稿之旅。

所谓初生牛犊不怕虎，加之把握不准文章质量（其实最开始的文章质量真不咋样的），所以“兴奋”之下，看着很多坛友介绍的经验，直接先投高级别的期刊，被拒了然后再依次往下。所以我第一篇文章（这篇文章不是农户生计方面的，是人口学方面的）先后投了 World Development, China Economic Review, Agriculture Economics, Population, Space and Place, Population and Environment 等期刊，然而要么是 3 天，要么 7 天，最多 20 天，都陆续的收到“Decision on Manuscript”，然后均是 I’m sorry to tell you that ……。还好中文期刊投稿有过被拒的经验，所以心里也不是很难受。然而陆续投了这么多期刊，还未送审就被枪毙在主编手里了，心里面还真有点不是滋味。我也知道第一次写英文文章肯定存在这样那样的不足，然而给我一个机会也好啊（送审给意见，被拒也行。毕竟有意见就可以更好的对文章进行改进）。本想慢慢的放弃去投个会议论文的，抱着最后一试的心态投了 Population Research and Policy Review，

然后网站状态一直是 under review，因为前面投稿一些期刊在主编手里也是 under review，所以我也不抱太大希望，结果一等就是一个多月。突然有天登录邮箱，标题 Decision on Manuscript，当时心又凉了半截。还好没有直接删除邮件，打开一看，结果是文章送审了 3 个审稿专家才被拒的。其中，一个是直接拒稿，觉得文章没多大意思，一个是大修，一个是倾向于接收。但总体意见中，有一条致命的问题是审稿人认为文章要做稳健性检验，自变量可能含有内生性问题。同时，正如前面所说，文章文献综述部分比较“散乱”，背后没有理论支撑（Theoretical Framework），未做研究假设等等。所以，主编考虑再三后，把文章给拒了。

虽然文章被拒，但那时我的心里是多么高兴啊。投稿这么多期刊，前后费了 3-4 个月时间，一直在被拒，连个意见都没有。现在虽然文章再次被拒了，但总好有详细的审稿意见。由于从写作到投稿，很多事情老师、课题组的师兄师姐都帮不了太多的忙（大家方向不一样，整个课题组也没发几篇 SCI/SSCI），所以大多数时候都是自己在这摸索。所以特别感谢 Population Research and Policy Review 这个杂志快速的给我审稿，给了详细的审稿意见（1 个多月时间 3 个审稿人审稿，不可谓不快）。让我的文章有了进一步修改的方向。于是，我的第一篇文章进入了修改的过程。

我个人感觉修改文章真的很难，比写作还难。写作时，有时可以天马行空

一点，没有那么多的约束。但改文章就被限定到一定范围内了，尤其是当审稿人提出我听都没听说过的一些概念（如当时不知道稳健性检验、内生性是什么东西，因为我不是学计量出身的，用的软件是 SPSS，而非 Stata,Eviews,Sas 等），那更加头大了。所以现在想想当时修稿真的很痛苦。不过还是硬着头皮改下来了。想着这个期刊有个审稿人是倾向于接收的，所以抱着试试的心态，和导师商量后将文章重投到 Population Research and Policy Review，结果 3 天左右，主编回复说非常遗憾，杂志决定不要的文章，即使修改后也不会再要了。虽然有点淡淡的失落，但还是接受了这个现实，找了下一家期刊（一个 4 区的 SCI，名字就不透露了）。然后经过 4 个多月漫长的等待，给了我一个大修（其实审稿人是倾向于接收的，修改意见大部分都是增添内容，如数据来源介绍还不够详尽，结果分析部分没有和其它同类研究做深入对比分析等）,修回去后现在还在返回给审稿人了，相信最近就会有结果。问题应该不会太大。

在这篇文章投到最后这个杂志审稿期间，通过看文献，结合自己对农村的感性认识,加之寒假回家好好的恶补自学了一下计量经济学（看的是伍德里奇《计量经济学导论》，我还写了一篇帖子叫“一个外行的计量经济学学习之路”，貌似也挺火，引起了很多坛友的共鸣，后来担任经管代码库实习版主，写了一个新帖“〖"万万"个论坛币〗计量实证的“摸底”考试”，也引起了很多坛友的兴趣。所以第二篇英文文章的写作就没有那么艰难了。

由于这篇文章是用的去年 8 月份在三峡库区调研的农户微观实证数据，是关于农户生计方面的。项目是我一个师兄在负责，所以文章挂了他为二作，写好后他也给我提了一些具有参考的意见。同时，恰逢他在国外访学一年，访学的 boss 也是做这方面研究的，所以他和我商量了一下，请那边 boss 帮我看看文章，提一些很好的修改建议，并对文章文字进行了大幅度的修改（由于国外的那个老师比较忙，所以我文章他第一次改了 1 个多月，我师兄帮忙催了一下才改好给我，很认真，他把文章打印出来边看边改的，做了很多批注，并提了很多有用的修改建议。我拿到后快速的对他提出的问题进行了修改，并返回给他，然后他又耗时 1 个多月，帮我把文章文字好好的修改了一下，可以说文章经他手一改，基本已经面目全非了。借此平台，感谢他无私的奉献，虽然后来挂了他为三作）。其中，记忆比较深刻的是他提到的多重共线性问题和极端异常值的问题。因为，最开始写文章时没处理极端异常值，导致描述性统计分析时，有 2 个左右指标标准差远远大于均值，这样肯定是不行的。从数据处理到文章写作，也是 3 周左右。再到前后师兄和国外的老师提意见修改，前后 3 个多月，终于把文章终稿弄好。然后在国外老师建议下再次投稿 World Development,结果几天后再次接到了拒稿通知。不过这次拒稿是说我文章研究区太小，可能吸引的读着有限。后来又投了 Journal of Development Studies,又被拒了，说文章不适合本刊，还建议投 Agriculture Economics 等农业方面的综合期刊。文章接连被拒，弄得我自己都

没啥信心了。所以后来看到本科导师在一个 SCI 和 SSCI 双检索的期刊上发了一篇文章（SCI 算 4 区，从投稿到发表 2 个多月），所以打电话咨询了一下他。我急切的把这篇文章投到这个期刊去了（期刊名这里就不透漏了）。因为当时文章一直在被拒，没多大信心，心里面对文章的质量也还是没底。所以想着，能发一个就不错吧，先达到毕业条件。那时，两篇英文文章加起来已经被拒了 8-9 次了，不由得不怕。结果一等又是 3 个月左右。

后来接到修稿通知，文章要大修，但还好没拒稿，要求 15 天内修回去。同时，我第一篇稿子也接到修稿通知（那时都想哭了：一是终于都没被再次拒稿了，修稿就表示有希望；二是两篇稿子都等着修，还催的比较急，那时还在帮导师写基金，感觉时间不够用啊，不过还是硬着头皮上了）。这里要说一下，我第二篇文章推荐了几个 top 期刊的审稿人，其中一个审稿人应该同意审稿了，给我提了很详尽的审稿意见，密密麻麻的 4 页纸，几十个问题。当时看到就有些崩溃，不过还是硬着头皮上了。这篇文章大修返回后 10 天左右又接到修稿通知了，不过这次是小修，要求 2 天内返回去。我快速的修好后，返回去 1-2 天后就接到收稿通知了。生平第一篇英文文章就给了这个 SCI 和 SSCI 双检索的期刊，从写作到发表前后耗时 8 个月左右（虽然从投稿到接收就 4 个多月）。

此外，本着再接再厉的原则，期间还写了一篇文章，经过一个多月初审，现在正在找审稿人，也不知道后续会怎样。前面啰啰嗦嗦说了这么多，大致是自己

的一个投稿历程和投稿过程中的一点心态变化。下面综合来说说我个人的一点从写作到投稿的一点经验。

第一，英语好的尽量还是下笔就英语，可模仿着 top 期刊上一些文章的句式写，但不要照抄，有些期刊要查重的，尽量写成自己的话。当然，如果有平台，如找国外同行的专家帮忙改文字就更好了。至少在语言关文章是过了。SSCI 是很讲究语言的表达的，从我前后两次修稿的过程就可看出，国外专家要抠字眼的，他只要有一个单词看不明白，就会问的。虽然有时以中国人的思维/常识来看，他/她提的问题就不是问题，但确实东西方文化和认知的差异客观存在。虽然我一直是半中半英，整体写好后再翻译，最后不行再拿到 AJE 润色（如果有导师支持这笔开支，当然无所谓。如果没有人支柱，完全靠自己，那就只有硬着头皮上了，毕竟润色费比较贵，而写 SSCI 的文章动不动就是 10000 字以上了）。

第二，文章的整体结构要心里有谱。这里我还是结合做的微观实证研究（应用计量经济学）来说。在我这个学科，写 SCI/SSCI 大致框架分为以下几个部分（这部分其实也可看看伍德里奇《计量经济学导论》中“一个经验项目的实施”这章，写得真的很经典）:

1.题目：题目要有吸引力。尽量还是不要写对 XX 的影响因素研究了（如对劳动力迁移区位的影响因素研究），这个影响因素就太多了。可以写你主要关注的因素，或者你研究结果与其它同类研究存在明显差异的因素。如同样是写劳动力

迁移区位的影响因素研究，我可以写基于推拉理论的劳动力迁移区位影响因素研究，基于性别差异的劳动力迁移区位影响因素研究，劳动力社会资本/社会关系网络对动力迁移区位影响研究等。突出理论视角，突出关注核心（虽然还是同样的自变量和因变量，但有的是核心变量，有的只是做了控制变量而已）。

2.摘要：英文期刊摘要有的要求是 100 字（如 World Development），但一般是 300-500 字。所以一般就 4 句话。第一句交代一下学术上的意义（我看很多期刊文章的第一句都是，关于 XX 方面的研究很多，而关于本文这方面的研究比较少），第二句交代利用什么数据和方法做了什么，第三句交代研究结果。第四句交代政策建议/研究启示等。其中，值得一提的是研究结果的交代，尽量不要把文章分析的结果全部罗列在这个地方，这也是我被收录的这篇稿子一个审稿专家提的建议，他建议我将有吸引力的和其它研究不同或文章独创的一些结果列在这里，而不是把所有结果一股脑儿的堆砌在摘要，突不出重点。所以，我看到我结果里面有几个变量是在中国的研究区实际和国外的一些文献结果不同，同时年龄与因变量有个倒 U 型的关系，所以这条建议的修改我就是把研究结果的不同及年龄这个放到摘要了。此外，审稿人指出摘要的关键词尽量不要和题目一样，这点也是值得注意的。

3.引言：引言我一直认为是比较难写的，很多时候我也是先分析数据然后得到结果，把后面所有部分都写完了再来写的引言和摘要。其实，这种方式在现在

的我看来，可能有些不好。为什么不好呢？因为引言部分开始就没梳理好，可能文章想解决的关键科学问题就没梳理好，就有点为模型而模型的意味了，有点像在玩数字游戏。可能把握住了文章要解决的关键科学问题，文章的结构框架，后面的数据处理分析可能才能有的放矢。结合看国外一些好期刊上的文章，引言大多数都是问题导向的。一般在引言最后都会说本研究主要的解决的关键问题如下…。所以，综合来看，引言是一个引子，是为了引出文章拟解决的关键科学问题服务的。提出了关键科学问题，审稿人一看，就明白文章的学术意义/政策意义大概是怎样的了。如果有兴趣，可能会进一步往下看。这在无形中可能能增加审稿人审稿的概率。

4.文献综述。这部分是重中之重，也是不好写的部分。因为大千世界，文献那么多，每个人的经历都有限，怎么可能把所有同类研究读完并很好的整合在一起呢。如果按照我以上的逻辑，那文献综述是没法写的。但其实这部分把握住以下关键问题就好。一是国内外的学者在文章这方面的研究做了些什么，现在做到什么程度，最重要的是还有哪些方面做得不足。而本文又正好多多少少能弥补一些这方面的不足。其实，不管文献综述多少，一定要把自己文章的特色体现出来，即本文相比于其它同类研究，哪些方面是新的，是进步的。同时，如果研究背后有理论支撑，要把理论摆出来（如我做农户生计生计，它的 Theoretical Framework 就是英国国际发展署的可持续生计分析框架）。我第一篇文章就因为

没有摆理论出来，所以投了那么多期刊杂志，最终审稿的两个杂志审稿人都提到了这个问题。即看似作者很好的掌握了国内外相关研究进展，但貌似作者文章背后没有一个合理的理论去支撑文章。而作者要提了一大堆指标体系，然后构建模型去做分析。如果没有理论支撑，这些指标的选取就显得无意义了。以上大致是审稿人的意见。所以，我第一篇文章就加入了推拉理论/新迁移经济学理论，第二篇文章可持续生计分析框架作为我文章的理论框架基础。

5.研究设计：这部分交代研究区域、数据来源、研究假设等。其中，值得一提的是研究假设的提出可根据第 4 部分文献综述结合自身对研究区实际感性认识而来，要做到有理有据。同时，像我这种做微观实证研究的，研究区介绍和数据来源一定要交代清楚。即我是以什么标准来抽样的，我的样本框是怎样（即我是从什么样的一个总体中抽样的），如何保证我的样本能很好的代表我的总体。如果做的是问卷，怎样保证问卷的信度和效度。同时，我这篇文章涉及到的问卷大概问了哪些方面的问题，一份问卷大概要多久等等。交代的越详细越好，因为耐心负责任的审稿人会紧抓这些问题不放的（我两篇稿子都被提到了这些方面的问题）。

6.研究结果：微观实证研究大概分为描述性统计分析和计量经济模型分析两块。描述性统计分析一般报道 Min, Max, Mean, SD。然而很多非搞计量出身的人认为做这部分是没多大作用的。这部分在我看来主要有以下两个方面作用：一

是大概看看数据的变化趋势和研究假设是否相符；二是看看数据是否做了相应的处理（就像我被收录的那篇文章一样，最开始是没有处理极端异常值的，所以SD>>Mean，这明显是有问题的，因为后面构建的模型中，偏回归系数因为极端异常值的影响，方向都改变了）。同时，注意在做描述性统计分析时，不用每个指标都说一下（即不用说那么详细），只报道你感兴趣的指标或结果比较有意思的指标即可，很多指标一句话带过即可；然后是构建计量经济模型，构建计量经济模型可能遇到的一些问题可参见我这篇帖子：http://bbs.pinggu.org/thread-3641714-1-1.html，考虑变量内生性、异方差性、遗漏变量等问题。最主要的是在做研究结果分析时，一定要与同类研究做对比，找出异同，存在明显差异的要思考差异背后原因，并做详细阐述。这部分可能是审稿人比较关心的部分，也是审稿人觉得有趣与否的部分。我第一篇文章审稿人就觉得这部分是 very interesting 的，这样无形中可能会为文章加分。当然，这部分的一些讨论也可放到第 7 部分去，而在这大部分只简要报道研究结果即可，看个人写作习惯。

7.研究结论/政策启示/结论与讨论：这部分可简要概括前面实证部分得到的结果，然后着重指出本研究可能存在的不足和对将来研究的一些启示/政策建议/政策启示。

8．参考文献：强烈建议使用 Endnote 管理参考文献。方便，实用，快捷。

Web of science 数据库检索文献的学习及 Endnote 的学习视频可从以下网址找到并下载：http://www.thomsonscientific.com.cn/productsservices/（我是自学成才的，相信需要的坛友也能）。

第三，系统的学习一下计量经济学，并熟练掌握一门计量经济软件。虽然我个人用了将近 7 年的 SPSS。但若想真心的做微观实证研究/应用计量研究，强烈推荐 stata，它轻巧方便，在回归分析这块功能很强，且有 do 文档保存命令语句，即使文章投稿后过几个月来修改，直接改动 do 文档即可。

第四，养成看专著做笔记的好习惯，同时有 idea 了不要拖延，快速行动。有时想法是转瞬即逝的，所以好记性不如烂笔头，有想法第一时间记录，并快速的去实施，不然放在那里过了就过了。看中文专著可以帮助系统的梳理概念框架，会提供很多有益的东西。毕竟好的专著大多数是一些学者几年的心血（很多都是在博士论文基础上改的），不可谓没有启示。我被收录文章的灵感就来自看专著，其中一些研究假设的提出都依赖于专著中提到的一些观点和引用的一些文献。

第五，文章被拒是常事，摆正心态。同时，要戒骄戒躁。学术不是一天两天就能做好的，真的要做好坐冷板凳的准备。得失心不要太重，该来的自然就来了。

第六，有平台文章写好后多请同行的师兄师姐，老师批评提意见。毕竟个人见识有限，文章都是在修改中出来的。我第一篇文章投那么多期刊不中，没有拿出去交流可能是其中一个很重要的原因。所以学术不要“闭门造车”，要“拿出去”。

最后，祝大家多发文章吧。越努力，越幸运，与君共勉。

2.6

农林经济类文章发表经验：两个学报和一个期刊

原文地址：http://bbs.pinggu.org/thread-3573439-1-1.html

原文名称：农林经济类文章发表经验：两个学报和一个农业科学期刊

想着自己最开始投稿的时候经常到人大经济论坛和小木虫等地方查询借鉴相关经验，现在自己有几篇小文章了，和大家分享一点自己的投稿经历，希望能帮助到需要的朋友。

第一篇文章——《广东农业科学》。

所谓初生牛犊不怕虎，在北京研究生院上研一期间，利用导师手里的调研数据，写了一篇描述性统计分析的文章，那时完全没经验。第一次投了《农业经济问题》后，几天收到邮件交了 100 大洋审稿费，然后进入漫长的等待期。中间经过 1 个半月收稿，1 个半月初审，1 个半月再审，最后被拒。从投稿到被拒总共历时 5 个多月。由于是第一次投稿，没经验，也被打击了一下。故而在导师建议

下改投了一个《农村经济》，这个杂志不知道怎么回事，投过去后石沉大海了，网上说是1个月左右会有结果。我还邮寄了纸质版，至今都杳无音讯。焦急的等待了1个半月后，打了个电话问了一下，杂志社编辑说没消息就是不用，态度感觉也不是很好。然后心灰意冷下投了《广东农业科学》，然后又是焦急的等待。20天左右吧，直接发录用通知了。不过需要2600大洋版面费，文章将近10000字，页码9个word。由于是生平第一篇文章，心里没底，又想发但又感觉有点贵，不好意思向导师开口说发（我导师人很好，就像父亲一样对待我们学生），最终还是老师来问我（他是通讯作者，也收到邮件了）这个期刊是不是北大中文核心，是就发。我说是，然后就让我从他课题本里报。第一篇文章基本就这么发了。

由于这个杂志是个半月刊，见刊比较快。没等多久，就收到邮件，让修改文字。没什么经验，简单修改后返回去就发表了。总体感觉这篇文章虽经历一系列波折，但也算轻松的发表了，期间算是较为平顺。

PS：这个杂志编辑部老师人很好说话，就是版面费略高。如果需要文章毕业或者评职称，可以酌情发表。不过如果想要在学术圈子里继续深造的，还可以继续研究发文章的规律。

第二篇文章——《中国农业大学学报（自然科学版）》。

这次还是利用调研数据写了一篇实证分析文章，用了有序多分类logistic回归模型，那时刚好在北京研究生院学习这方面的统计知识，就想着学以致用。由

于第一篇文章的发表基本没什么经历可言，故而这篇文章写出来了又尴尬了，不知道投什么杂志。主要是不知道自己写出来的文章水平怎么样，导师也不是搞这方面的，也给不了什么建议。一切还是得靠自己摸索。思前想后，《中国农村经济》、《中国农村观察》这类顶级的期刊，自己水平肯定达不到。《农业经济问题》、《农业技术经济》这类期刊上篇文章投过，一来因为审稿周期长，二来因为被拒过，没啥信心也不敢投。最后自己琢磨下，找到了《中国农业大学学报（自然科学版）》这个杂志，抱着试试的态度投过去了。开始估计没戏，但又没收到通知，故而只有焦灼的等待了，基本上天天去系统里刷刷，看看状态有没有变。

漫长的等待终于等到了结果。一个月后，一个北京的陌生号码给我打电话。当时我在财务报账，本来不想接的（因为经常接到一些 400 开头的推销电话，比较烦），但还是接了。结果喜出望外，是编辑部的刘老师，她先确认我身份，然后说审稿结果回来了，需要修改，让我下午给她打电话，她给我说哪些地方需要改。

然后就是反复的修改过程。中间反复的修改了 3-4 次。第一次是主要修改审稿专家提的意见。每条逐一的推敲后，很有礼貌地逐一回答，并改正。然后几次都是修改文章的措辞。这里要给农大学报 32 个赞，大到每一句话，小到一个标点，一条参考文献。刘老师都事无巨细的给我指出要怎么改，使我受益匪浅最终使得文章得以顺利发表。文章 9 个版面，200 元/页版面费+100 元审稿费，最终 1700 元版面费顺利发表。

经过这么一个反复的过程，我自己学到了很多宝贵的经验。尤其是那段来回修改文字的过程让我受益匪浅。这也让我在以后文章写作过程中更注意表达的方式技巧，以及参考文献标注等每一个细小的细节。这篇文章算是我学术起步开始的之作了。再次感谢农大学报的老师。

如果是实证分析的文章，尤其是用了计量经济模型的农林经济类文章。对自己文章水平比较有信心的可考虑投这个期刊，拿得出手，重要的是审稿快，期刊质量也可以，版面费也还可以接受。

PS：农大学报的出版质量真的没得说，真的是很好很好。

第三篇文章——《西南大学学报——自然科学版》

有了农大学报的发表经验，感觉自己自信了不少。不久后又有了一篇农户生计的文章问世。经过写作和老师指出问题修改后，选取期刊又成了一个问题。由于有了农大学报在前，这次自己有点自信过头，将文章投到《中国农村经济》这个杂志去了，然后开始了漫长的等待之旅。1 个月，2 个月·6 个月，期间内心反复煎熬。又想撤稿，又有点不甘心。由于不知道结果，也没办法另投期刊，只有焦急的等着。等了 7 个月，实在等不住了，给那边编辑部打电话。编辑部老师态度很好，说 7 个月没消息应该就是不用了，审稿的主编老师那天不在，所以也没法继续问为什么不要我的文章，问问意见之类的。等待了 7 个月换来这么一个结果，内心有点失落，但同时又感觉解脱了。毕竟内心早就做好被拒准备。现在知

道结果了，正好可以光明正大的另选期刊投稿了。

本来想再投投农大学报试试，毕竟审稿周期短（1 个月左右给结果），但农大学报的表格要求除了数据之外，所有表格里的内容都要中英文对照。而我这篇文章表格 5-6 个，里面文字也比较多，实在是难改，所以思虑再三。未投这个期刊了。最终找了半天，投了这个《西南大学学报《自然科学版》，结果一等就是半年。

《西南大学学报》分自然科学版和社会科学版，社科版好像不要审稿费，也比较快（投稿之后才发现）。自然科学版投稿一周左右，编辑部发邮件来说收稿了，需要邮寄 100 大洋审稿费，收到审稿费后才会送外审专家。邮寄审稿费后，4 天后编辑部状态变为外审。外审一等就是 2 个月。实在等不住了，打个电话去问现在状态。编辑部老师人还不错，说审稿结果回来了，现在进入副主编审稿。挂了电话，几分钟后一查状态变了。然后一等又是将近两月，正准备打电话去问的那天，一查状态刚变为主编终审。然后又是漫长的等待，可能是因为年边了，主编是校长，比较忙，没时间看。故而这个阶段我等了又将近 2 月，期间打电话问了一下。说是基本上外审通过，就没多大问题，会被录用。副主编和主编这两个阶段更像是“拖延”时间感觉西南大学本校的老师或学生投稿这个期刊 3-4 个月就发表了，2 个月左右结果出来了。

导师和我都收到邮件，小修，修改后发表。有了农大学报的经验，这次好好

地按照审稿人的意见修改了文章（PS：意见还是很中肯，很专业）。然后返回去。然后是责任编辑阶段，一直持续到我写这篇稿子，一直都还是这个状态。期间我忍不住好奇打电话去问了，那边老师说现在在等着文章查重，查重没问题基本上就是等着排版，安排发表期数了。最后这个是我关心的问题，问了一下。编辑部老师说《西南大学学报《自然科学版》》从投稿到录用发表一般周期是 1 年左右，我的稿子是 2014 年 6 月 11 号投的，她说估计最早今年 7 月份见刊。我想，反正我硕博连读了，可以等等。就说行，希望早点见刊就行。现在在等最后让交版面费。

就这么一折腾，这篇文章从写出来到最终发表，算上《中国农村经济》的 7 个月，这里的 1 年，中间再折腾一下，都快 2 年了。这给我的经验教训就是在投稿前一定要想方设法的请专业人士给自己的文章把把关，推荐推荐期刊。不要一个人在那里瞎摸索。此外，选定一个期刊后，一定要去期刊查查这几年有没有发表和自己文章主题比较相关的文章，判定主题符不符合。甚至可以打电话给编辑部，问问收不收这方面的文章。然后，通过人大经济论坛、小木虫等地方找找有没有选定期刊的投稿经验。看看审稿周期，版面费之类的。不要随便选一个期刊就瞎投了。前期工作做得越足，后期等待就不会那么焦急。我这算是血淋淋的教训了（虽然最终被收录了）。

注：《西南大学学报（自然科学版）》这个期刊需要有国家级基金支持，同时

一作必须是博士生或者取得博士学历的人。门槛比较高。学生我真的是不建议投这个杂志了，比较时间太长了。从投稿到接收都基本要 8 个月，再到出版都 1 年多了。

以上就是我到目前为止投中文期刊的一点浅薄经历，详尽的贴出来，希望和农经的小伙伴交流。少走弯路，也希望能帮到需要帮助的人。

注：2014 年 9 月进入博士学习阶段，到目前为止，就没写过中文文章了。写了 3 篇英文，期间被拒 7-8 次，现在都还在审。等有结果了，我将自己英文投稿的经历也和大家分享交流一下。

这个帖子有点长。如果大家能耐心看完就非常感谢了。祝大家新的一年发更多的文章吧。另外，期刊投稿动态随时都在变化，本文的有些内容可能不太准确，大家可以随时关注期刊官方网站，了解最新动态，以便写作时把握，祝好运。

2.7

小议计量经济学检验方法

注：下文由胖胖小龟宝始发于经管之家（原人大经济论坛）

原文地址：http://bbs.pinggu.org/thread-3656363-1-1.html

计量经济学中的检验方法多种多样，而且在不同的假设前提之下，使用的检验统计量不同，在这里我论述几种比较常见的方法。

在讨论不同的检验之前，我们必须知道为什么要检验，到底检验什么？如果这个问题都不知道，那么我觉得我们很荒谬或者说是很模式化。检验的含义是要确实因果关系，计量经济学的核心是要说因果关系是怎么样的。那么如果两个东西之间没有什么因果联系，那么我们寻找的原因就不对。那么这样的结果是没有什么意义的，或者说是意义不大的。那么检验对于我们确认结果非常的重要，也是评价我们的结果是否拥有价值的关键因素。所以要做统计检验。

t 检验，t 检验主要是检验单个 ols 估计值或者说是参数估计值的显著性，什

么是显著性？也就是给定一个容忍程度，一个我们可以犯错误的限度，错误分为两类：1、本来是错的但是我们认为是对的。2、本来是对的我们认为是错的。统计的检验主要是针对第一种错误而言的。一般的计量经济学中的这个容忍程度是 5%，也就是说可以容忍我们范第一类错误的概率是 5%。这样说不准确，但是比较好理解。t-stastic 是类似标准正态化的正态分布两一样，也就是估计值减去假设值除以估计值得标准差，一般假设值是 0，这一点不难理解，如果是 0 ，那么也就意味着没有因果关系。这个 t-static 在经典假设之下服从 t 分布。t 分布一般是和正态分布差不多，尤其是当样本的量足够大的时候，一般的经验认为在样本数量大于 120 的时候，就可以看成是正态分布的。

F-statistc：F 检验是属于联合检验比较重要的一种，主要的目的是用于对于一系列的原因的是否会产生结果这样一个命题做出的检验。F 统计量主要的产生来源是 SSR\SST\SSE 三个量。但是这个检验有一个缺点是必须在经典假设之下才能有效。

LM 检验：这个检验的性质和 F 检验的性质是一样的，都是检验联合显著性的，不同的是 F 统计量符合 F 分布，但是 LM 统计量服从卡方分布。卡方分布是正态分布的变量的平方和，而 F 分布是卡方分布的商，并且分子和分布必须独立，这就是为什么 F 检验适用范围受限的原因。LM=n*SSR、或者是 LM=n-SSR。

至于其他的 White 检验、Brusch-pagan 检验（异方差的检验方法）、还有

序列相关的 t 检验、DW 检验基本原来是相同的。

关于异方差检验、序列相关的检验其中存在不同的地方，但是思想基本是相同的。

关于异方差检验的讨论：

1、Brusch-pagan 检验：这个检验的思路比较简单，主要是要研究残查和 X 之间的关系，给定这样的一个方程：u=b0+b1*x1+……+bn*xn+u'的回归，其中进行 F 检验和 LM 检验。如果检验通过那么不存在异方差，如果不通过那么存在异方差。

2、White 检验：这个检验也是对异方差的检验，但是这个检验不同的是不仅对于 X 的一次方进行回归 ,而且考虑到残查和 x 的平方还有 Xi*Xj 之间的关系。给定如下方程：u=b0+b1*y+b2*y^2+u'。也是用 F 和 LM 联合检验来检验显著性。如果通过那么不存在异方差，否则存在。

序列相关的检验方法的讨论：

对于时间序列的问题需要知道一个东西，也就是一阶自回归过程，也就是一般在教科书中说到的：AR(1)过程，其中的道理主要是说在当期的变量主要是取决于过去一个时期的变量和一个随机误差项。表示如下：Ut=p*U(t-1)+et。在这里我要说到几个概念问题，I(1)（一阶积整）、I(0)（零阶积整）。其中的一介自回归过程 AR(1)就属于零阶积整过程，而一阶积整过程实际上是随机游动和飘移的

随机游动过程。随机游动过程：Ut=U(t-1)+et。也就是在 AR(1)的过程之下，其中的 P 是等于 1 的。飘移的随机游动过程：Ut=a+U(t-1)+et。其中随机游动过程和 AR(1)过程中的不同点在于一个弱相依性的强弱问题，实际上我们在时间序列问题中，我们可以认为任何一个过程是弱相依的，但是问题的关键是我们不知道到底有多弱？或者更加直观地说，我们想知道 P 到底是多大，如果 P 是 0.9 或者是一个比较接近于 1 得数 那么可能我们可以认为这个时间序列有高度持久性，这个概念表示当期的变量却绝于一个很早的时期的变量，比如一阶积整过程，实际上 et 是一个独立统分布的变量，而且条件数学期望等于 0，没有异方差性。那么实际上这个序列的数学期望是和期数没有什么关系的。那么也就意味着从第 0 期开始，U 的数学期望值就是和很久以后的 U 的数学期望值一样的。但是方差就不同了，方差随着时间的增加不断扩大。我们知道了，这种不同的概念就可以讨论在一阶自回归的条件之下的检验问题，但是我们说一介自回归的过程是参差序列的特征而已，其他的变量的特征问题我们不谈。

在讨论检验的问题以前，我有必要交待一下时间序列在 ols 估计的时候我们应该注意什么。实际上解决序列自相关问题最主要的问题就是一个差分的方法。因为如果是长期持久的序列或者是不是长期持久的序列，那么一定的差分就可以解除这种问题。

2.8

学习心得与汇总之 STATA 基本操作汇总：异方差、序列相关、多重共线性

注：下文由 crystal8832 始发于经管之家(原人大经济论坛)

原文地址：http://bbs.pinggu.org/thread-3711175-1-1.html

自从蓝色版主（人大经济论坛资深版主）对 STATA 版块做了较为细致的汇总之后，已经很久没有人对其进行进一步的完善了。本人最近准备重新对版块的内容进行梳理，把坛友们经常询问的问题再一次进行汇总，以方便大家的使用。今天主要是对计量的基本问题的相关汇总。

一、异方差问题

STATA 检查是否存在异方差的方法：1.看残差图【 模型回归之后使用即可】

rvfplot（残差与拟合值的散点图）

rvpplot（残差与解释变量的的散点图）

2.怀特(White,1980）检验【 模型回归之后使用即可】

estat imtest,white（怀特检验）whitetst（外源程序，需下载）

3.BP(Breusch and Pagan,1979)检验【 模型回归之后使用即可】

estat hettest（默认设置使用拟合值 y_hat）

estat hettest（使用方程邮编的解释变量，而不是 y_hat）

estat hettest varlist（指定使用某些解释变量）

解决办法：

1.WLS 加权最小二乘法

reg y x1 x2 x3 [aw=1/var]

eg: reg y x1 x2 x3

predict e1,res

g e2=e1^2

g lne2=log(e2)

reg lne2 y,noc

predict lne2f

g e2f=exp(lne2f)

reg y x1 x2 x3 [aw=1/e2f]

2.White（1980）

eg: reg y x1 x2 x3,robust3. wls0 命令

相关帖子传送：

http://bbs.pinggu.org/thread-276344-1-1.html

http://bbs.pinggu.org/thread-2999151-1-1.html

http://bbs.pinggu.org/thread-418802-1-1.html

二、序列相关问题

STATA 检查是否存在序列相关的方法：

1.画图

在做完回归之后，先生成残差项 e

scatter e L.e

2.BG 检验

estat bgodfrey（默认滞后阶数为 1）

3.Ljung-Box Q 检验

eg: reg y x1 x2 x3

predict e,res

wntestq e

3.DW 检验

estat dwatson 解决办法：

1.Newey 稳健性标准差

newey y x,lag(p) (滞后阶数必选)

2.可行广义最小二乘法（FGLS）

prais y x

prais y x,corc

相关帖子传送：

http://bbs.pinggu.org/thread-3035976-1-1.html

http://bbs.pinggu.org/thread-2384962-1-1.html

三、多重共线问题

多重共线性并不会改变 OLS 估计量 BULE 的性质，但会使得对系数的估计变得不准确。

STATA 检查是否存在多重共线的方法：

estat vif

VIF 值越大说明多重共线性问题越严重。一般认为，最大的 VIF 不超过 10，则不存在明显的多重共线性。

解决办法：

1.如果只关心方程的预测能力，则在整个方程显著的条件下，可以不必关心具体的回归系数。

2.增加样本容量，剔除导致多重共线性的变量或者修改模型设定形式。

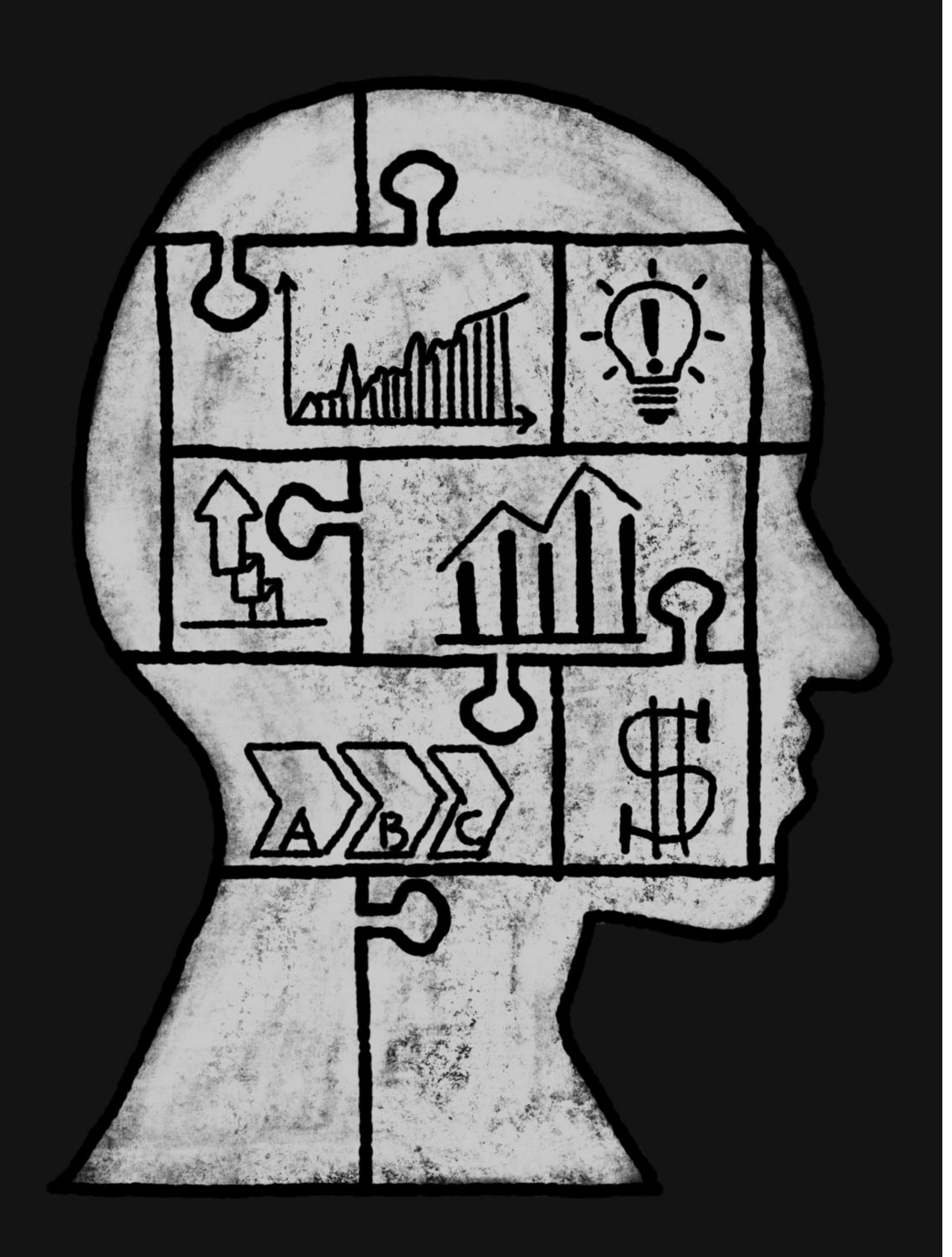
3.对于时间序列样本，通过使用差分模型可以一定程度上消除原模型中的多重共线性。

4.岭回归方法。

相关帖子传送：

http://bbs.pinggu.org/thread-1263473-1-1.html

http://bbs.pinggu.org/thread-1473811-1-1.html



拨开计量与统计的面纱

三、

拨开计量与统计的面纱：软件学习精彩答疑与学习资源荟萃

〖我在写第一篇SCI时由于没有系统学习过计量的东西，只是使用 SPSS 来实现自己的想法，这导致文章在实证模型上就弱了一些。我慢慢地发现 SPSS 解决很多问题似乎不够用了，且在书中反复出现 Eviews，STATA 等统计分析软件的字眼。这推着我往前思考，是不是在学习 SPSS 软件的基础上要再另外系统地学习一门统计分析软件。SAS，我在研究生院上课时有过接触，是很强大，然而感觉比较难学，同时安装都起来都感觉比较麻烦。于是我慢慢地缩小范围，决定在 Eviews 和 STATA 两个软件里选一种软件学习。〗

3.1

软件学习方法



提问：

实证研究零基础博士生请教怎样选择、运用模型进行数据分析，看什么书可以入门？

我是农业经济管理专业博一的学生，马上面临开题的问题，但是以前读研时从没做过实证方面的论文，论文全部是文字性的，关于模型的选择、建立以及使用一窍不通，导师又要求博士论文中必须运用模型进行数据分析，发表的小论文中也大多会偏重于这方面，可自己这方面什么都不会，完全没有概念，现在很着急，请问各位大侠有什么书籍可以入门的吗？或者是大家介绍一下是怎么从不懂到懂的学习方法？最近因为开题以及发表小论文的问题，弄得自己非常焦虑，总想着要退学哎，现在想试试到论坛里求助一下过来人，小女子谢过了！

解 答 栏

计量经济学学习辅之以一门软件。鉴于楼主的基础，建议古扎拉蒂《计量经济学基础》和伍德里奇《计量经济学导论》，这两本书帮你掌握原理。软件建议 Eviews 或者 STATA，前者注重窗口操作，后者注重编码，我用的是 STATA。STATA 书籍推荐《高级计量经济学及 STATA 应用》和《用 STATA 学微观计量经济学》，这两本书配合前面的两本原理教材看，然后懂 STATA 操作。要解决农经类的一般问题还是没问题的。PS：我现在也是博一，计量的东西完全自学过来的。祝好运~



提问：

想学习数据分析，毫无基础，应该从哪些软件学起？

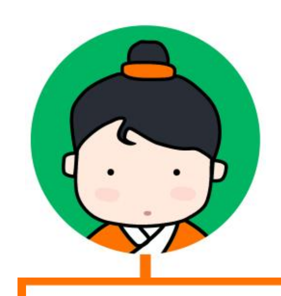
想学习数据分析，毫无基础，应该从哪些软件哪些参考书开始？求大神指点路径。

解 答 栏

推荐楼主看看我这个外行自学统计/计量的两个经验贴——一个外行的计量经济学学习之路，应该对你有启发。链接如下：

http://bbs.pinggu.org/thread-3596375-1-1.html；

http://bbs.pinggu.org/thread-3880985-1-1.html。祝好运~



提问：

正在看张文彤的《SPSS 统计分析高级教程》，看不懂怎么破？

正在看张文彤的《SPSS 统计分析高级教程》，看不懂怎么破？看到线性混合模型就很吃力了，感觉设定好复杂，而且好多不知道为什么要这么设定。

等看到广义线性模型、广义估计方程这一块，彻底崩溃了！完全不知道在说什么！

这要怎么办？还有哪些SPSS方面的书推荐吗？或者其他统计学的书推荐一下？万分感谢！

解 答 栏

线性混合模型、广义线性模型和广义估计方程这些内容应该属于相对“高级”，平时大家用的可能相对较少的内容。这些模型背后可能包含很多隐含的假设，而这些假设可能需要一定的数学功底或者计量经济学基础才能很好的理解。故而，楼主你直接凭借张文彤老师的《SPSS 统计分析教程》去自学，自然会比

较吃力。鉴于此，建议有二：

一是多学一些常用的统计分析方法（如 OLS，Logistic 回归，因子分析等），诸如广义线性模型这种相对“冷”门的方法需要用时再来系统的学习吧。

二是如果真的需要用到广义线性模型，建议先去学一些计量的基础知识，打好一定的基础后再来理解这个模型，估计就要好一些了。祝好运~



提问：

SPSS 在数学建模里有什么主要的作用？

数学建模中经常用 SPSS 来做什么？求大神来指导一下。马上就要做数学建模美赛了，希望学一些 SPSS 来增强实力！另外 SPSS 在处理大数据（数据量在 2 万——4 万，主要做数据的整合，数据筛选与多元回归分析）上与 r 相比，哪一个比较好一些；因为我们专业课有学过一些 r，然后 r 毕竟是编程，虽然更灵活一些，但是总觉得会比较麻烦；请您指导一下，谢谢！

解 答 栏

楼主这个问题的逻辑可能搞反了。我们在做一项研究时，一般先有要解决的问题，然后才会去寻求用什么手段（如用什么软件）来解决这个问题。故而，问题的逻辑不是 SPSS 能做什么，而是我们需要实现什么，SPSS 能否实现我们想要做的。

我没有用过 R，如果数据量比较大的话，还是建议用 R 吧。毕竟感觉现在 R 发展很迅速，应该有其独特之处。而且你要做的（多元回归）也挺简单的，用 R 几句代码就实现了吧。祝好运~

3.2

SPSS

3.2.1 数据排序（Sort Case）

操作说明如下

选择菜单【数据】→【排列个案】，打开对话框，如图 3.1 所示。

3.2.2 抽样（Select Case）

在统计分析中，有时不需要对所有的观测进行分析，而可能只对某些特定的对象有兴趣。利用 SPSS 的 Select Case 命令可以实现这种样本筛选的功能。以 SPSS 安装配套数据文件 Growth study.sav 为例，选择年龄大于 10 的观测，基本操作说明。



图 3.1 排列个案对话框

如下：

打开数据文件 Growth study.sav，选择【数据】→【选择个案】命令，打开对话框，如图 3.2：

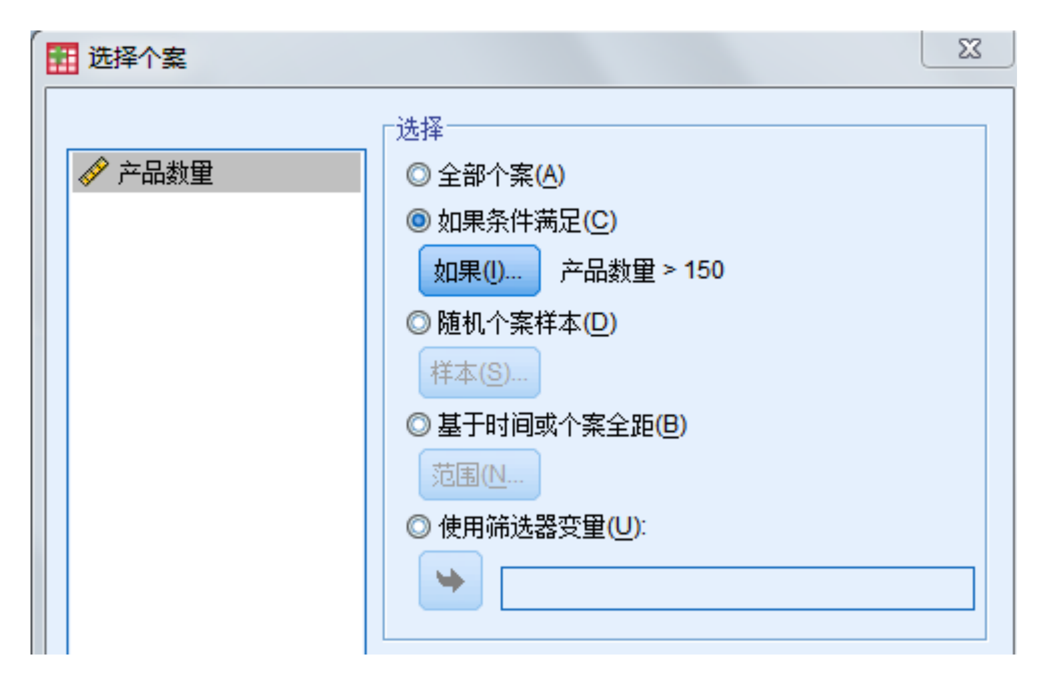


图 3.2 选择个案对话框

指定抽样的方式：【全部个案】不进行筛选；【如果条件满足】按指定条件进行筛选。本例设置：产品数量>150，如图 3.3 所示；

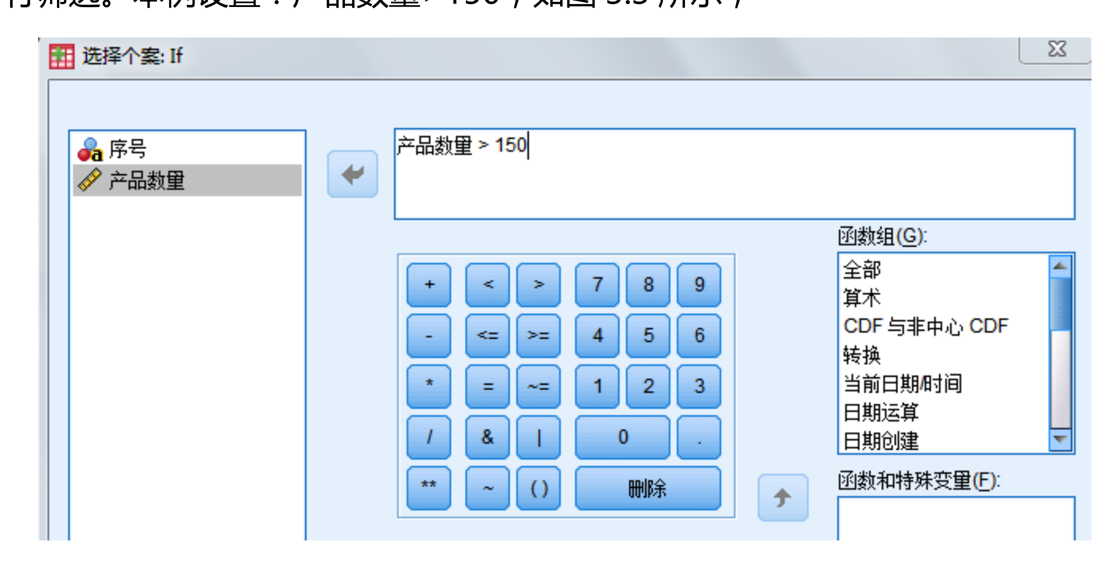


图 3.3 选择个案对话框

设置完成以后，点击 continue，进入下一步。

确定未被选择的观测的处理方法 这里选择默认选项【过滤掉未选定的个案】。

输出
过滤掉未选定的个案(F)
将选定个案复制到新数据集(O)
数据集名称(S):
删除未选定个案(L)

单击 ok 进行筛选，结果如图 3.4

	序号	产品数量	filter_$
1	1	160.00	1
2	10	150.00	0
3	11	162.00	1
4	12	156.00	1
5	13	179.00	1
6	14	179.00	1
7	15	151.00	1
8	16	157.00	1
9	17	154.00	1
10	18	179.00	1
11	19	148.00	0
12	2	170.00	1
13	20	156.00	1
14	3	181.00	1
15	4	156.00	1
16	5	176.00	1
17	6	148.00	0
18	7	198.00	1
19	8	179.00	1
20	9	162.00	1

图 3.4 选择个案的结果

3.2.3 增加个案的数据合并（【合并文件】→【添加个案】）

将新数据文件中的观测合并到原数据文件中，在 SPSS 中实现数据文件纵向合并的方法如下：

选择菜单【数据】→【合并文件】→【添加个案】，如图 3.5，选择需要追加的数据文件，单击打开按钮，弹出 Add Cases 对话框，如图 3.6。

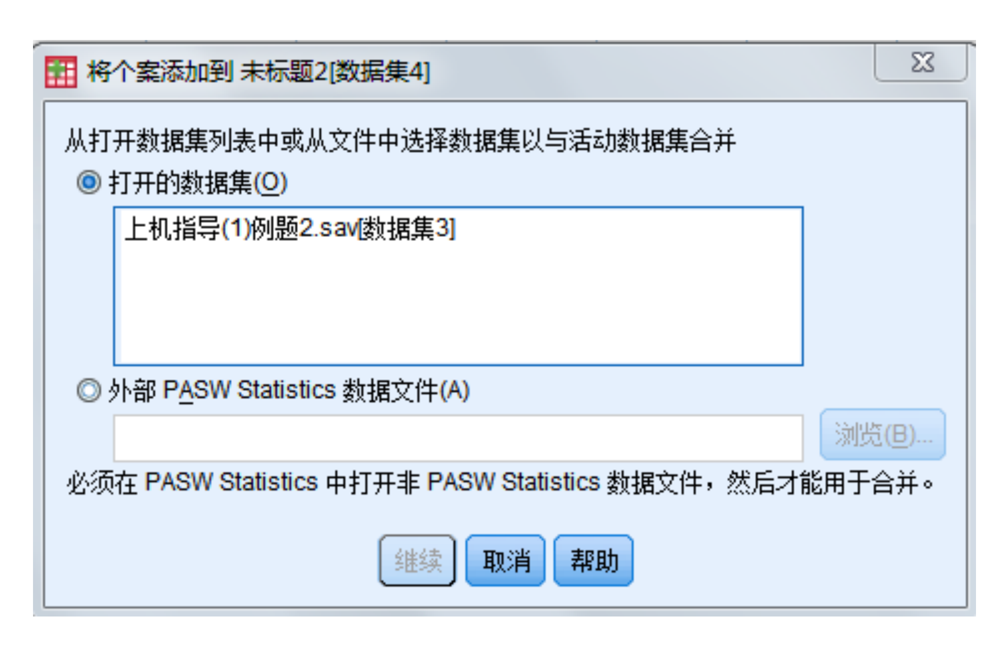


图 3.5 选择个体数据来源的文件

图 3.6

3.2.4 增加变量的数据合并（【合并文件】→【添加变量】）

增加变量时指把两个或多个数据文件实现横向对接。例如将不同课程的成绩

文件进行合并，收集来的数据被放置在一个新的数据文件中。在 SPSS 中实现数据文件横向合并的方法如下：

选择菜单【数据】→【合并文件】→【添加变量】,选择合并的数据文件，单击“打开”，弹出添加变量，如图 3.7 所示。

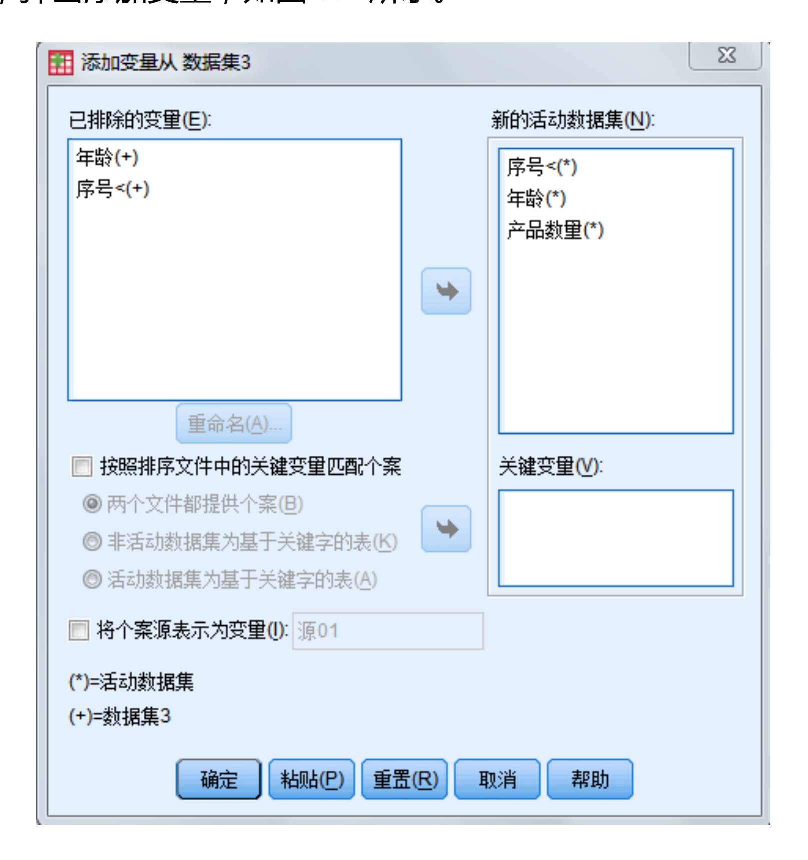


图 3.7

单击 Ok 执行合并命令。这样，两个数据文件将按观测的顺序一对一地横向合并。

3.2.5 数据拆分（Split File）

在进行统计分析时，经常要对文件中的观测进行分组，然后按组分别进行分

析。例如要求按性别不同分组。在 SPSS 中具体操作如下：

选择菜单【数据】→【分割文件】，打开对话框，如图 3.8 所示。

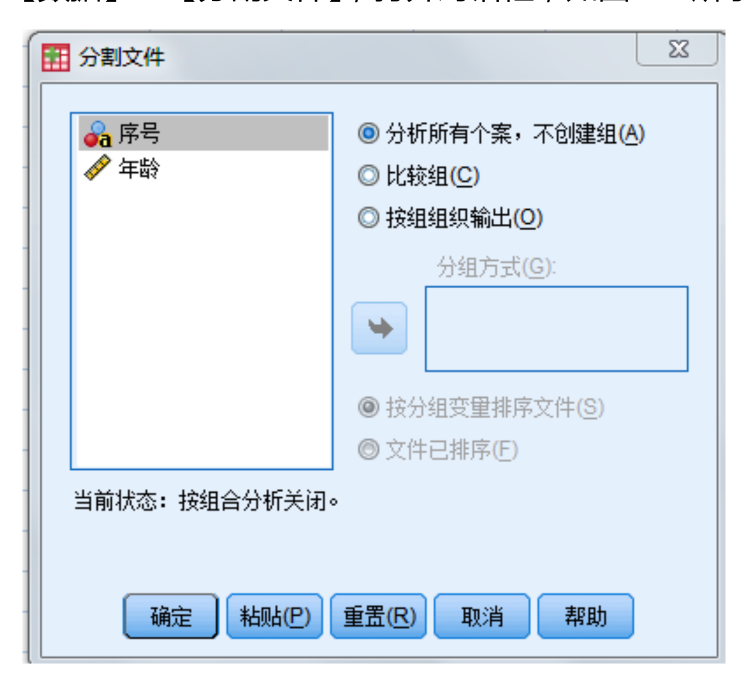


图 3.8 分割文件对话框

选择拆分数据后，输出结果的排列方式，该对话框提供了 3 种方式：对全部观测进行分析，不进行拆分；在输出结果种将各组的分析结果放在一起进行比较；按组排列输出结果，即单独显示每一分组的分析结果。

选择分组变量

选择数据的排序方式

单击 ok 按钮，执行操作

3.2.6 计算新变量

在对数据文件中的数据进行统计分析的过程中，为了更有效地处理数据和反

映事务的本质，有时需要对数据文件中的变量加工产生新的变量。比如经常需要把几个变量加总或取加权平均数，SPSS 中通过【计算】菜单命令来产生这样的新变量，其步骤如下：

选择菜单【转换】→【计算变量】，打开对话框，如图 3.9 所示。



图 3.9 Compute Variable 对话框

在目标变量输入框中输入生成的新变量的变量名。单击输入框下面类型与标签按钮，在跳出的对话框中可以对新变量的类型和标签进行设置。

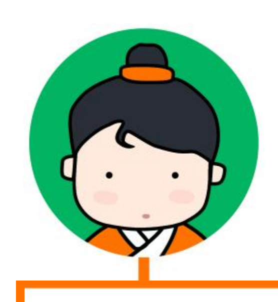
在数字表达式输入框中输入新变量的计算表达式。例如“年龄>20”。

单击【如果】按钮，弹出子对话框，如图 3.10 所示。包含所有个体：对所有的观测进行计算；如果个案满足条件则包括：仅对满足条件的观测进行计算。

单击 Ok 按钮，执行命令，则可以在数据文件中看到一个新生成的变量。



图 3.10 如果…子对话框



提问：

应该如何对连续变量进行设置？

我正在用广义估计方程做一个交通事故预测，其中的变量如上图，里面有“曲率变化率”、“坡度累计效应”这样的连续变量，也有“避险车道”这样的二项变量。但是模型跑出来之后，就都成离散变量了，请问应该如何对连续变量进行设置？

路段最大曲率km	弯坡组合km	坡度累计效应	直线接平曲线效应	避险车道	肇事货车比例
.526315789474	-.060057471264	.5	0	0	.8
.526315789474	.325000000000	1.0	0	0	.5
.384615384615	.114942528736	1.3	0	1	.3
.384615384615	.425287356322	1.1	0	0	.2
1.315789473684	-.460000000000	1.5	0	0	.4
1.315789473684	-.130574712644	1.1	0	1	.4
.625000000000	-.204022988506	1.3	0	0	.6
.909090909091	-.087643678161	1.4	0	1	.7
.952380952381	-.008189655172	1.5	0	1	1.0
.952380952381	-.341522988506	1.9	0	0	.5
.0	.287586206897	2.2	0	0	.9
1.388888888889	-.337816091954	2.4	1	0	.8
1.388888888889	-.148160919540	2.6	0	0	.7
1.333333333333	-.462500000000	3.5	0	0	.9
1.612903225806	-.207643678161	4.0	2	0	.9
1.666666666667	-.373563218391	4.5	0	0	1.0
1.666666666667	-.212643678161	4.9	0	1	1.0

解 答 栏

出现这种情况可能的原因是楼主在“变量视图”中将变量的属性设置错了，即本来应该是“度量”尺度的变量，楼主将其设置为“名义”或“有序”的了。可去检查一下变量设置情况。祝好运~



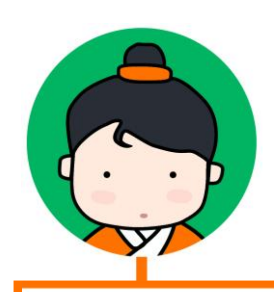
提问：

请问 SPSS 中导入的数据中默认都是字符串，改成数值型

改成数值型后为什么有些数据缺失了，本来有数的，变成空白了，应该怎么处理呢。

解 答 栏

没有遇到过这种情况，不清楚楼主具体是哪一步操作出现了问题。我有一个解决这个问题的“笨”办法，即不通过导入数据的形式将数据导入到 SPSS 中，而是直接将 Excel 中的数据复制到 SPSS“数据视图”中，然后再到“变量视图”设定变量名称和变量类型。祝好运~



提问：

SPSS 中计算变量生产新变量的还原问题

请教大家一个问题，我用 SPSS 软件根据几个变量用“计算变量”方法计算了一个新变量，但现在忘了当时的计算公式，不知 SPSS 软件中可以查看新变量的计算过程吗？如何查看？望各位大侠不惜赐教，谢谢！

解 答 栏

这个据我个人所知是不能查询的，所以建议楼主还是仔细回想一下当时的具体计算过程吧。祝好运~



提问：

SPSS 怎样在日期变量的原始数据中取年度值？

SPSS 怎样在日期变量的原始数据中取年度值？

解 答 栏

据我个人所知，SPSS 的窗口操作应该解决不了楼主的这个问题，不过可以编写代码试试。SPSS 我没试过编写代码，但 STATA 可以提取。PS：如果数据量不大，楼主可以先对数据进行排序，然后自己录入一个年份，再进行复制粘贴。方法虽然笨了点，但总能实现的。祝好运~



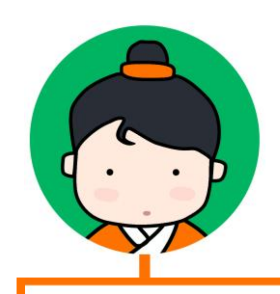
提问：

怎样将一个变量分为两个？

如何将 SPSS 中一个变量变成两个，例如价格收入这个变量变为女性收入，SPSS 中性别这个变量也给出了。高手帮回答下吧。

解 答 栏

楼主是要求女性价格收入的均值还是要得到一个新的变量，该变量是女性的价格收入呢？如果是第一种，直接使用描述统计分析即可；第二种可将价格按性别排序，然后直接将女性的价格部分复制出来即可，或者使用 recode 命令，另存为新变量实现。祝好运~



提问：

求推荐从建模到结论的书籍，附带 SPSS 或 SAS 程序

人统计小白一枚，开始从统计理论学起，但始终觉得不实践记不牢，学过就忘，都要没信心了。跪求权威教程，可以边学理论边通过程序操作的那种，万分感谢！

解 答 栏

SPSS 方面推荐张文彤老师《SPSS 统计分析教程》和《IBM SPSS 数据分析与挖掘实战案例精粹》，尤其是后面一本书，是以解决实际问题为导向而撰写的，从问题的梳理，到可以用的统计分析方法都给了明确的示范操作，看后应该会受益匪浅的。

至于想把实证（计量）学好，推荐伍德里奇《计量经济学导论》，从根本上把原理掌握。同时，辅之学习 STATA/EVIEWS/SAS 等软件，把计量原理用到实战上去。推荐楼主看看我这个外行自学统计/计量的两个经验贴——一个外行的计

量经济学学习之路，应该对你有启发。链接如下：

http://bbs.pinggu.org/thread-3596375-1-1.html；

http://bbs.pinggu.org/thread-3880985-1-1.html。祝好运~



提问：

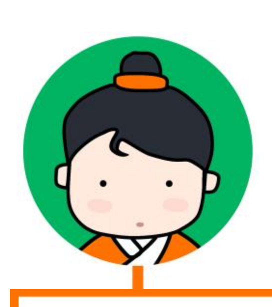
初学者怎么学习 AMOS 软件啊？

各位论坛大神啊，请问对于初学者，怎么入门学习 AMOS 啊？学习 AMOS 要有哪些预备知识呢？已经下载了该软件，可是界面操作都不是很熟。

解 答 栏

在论坛上下载或者买吴明隆老师《结构方程模型——AMOS 的操作与应用》一书（链接：http://bbs.pinggu.org/thread-965431-1-1.html），我就是以这

边书为蓝本自学的，并在这本书的指导下完成了结构方程模型的实证分析。这本书的优点就在于它以图示的方式系统的介绍 AMOS 软件的使用，是非常好的入门教材。祝好运~



提问：

关于 SPSS 自由度问题？？？

KMO 和 Bartlett 的检验时候。

取样足够度的 Kaiser-Meyer-Olkin 度量。		.799
Bartlett 的球形度检验	近似卡方	99.783
	df	12
	Sig.	.000

自由度只有 12，这样合理吗？

解 答 栏

合理的，结果没问题，可以做主成分分析/因子分析。建议楼主找本教程看看自由度的计算是怎样的，把这个搞清楚了就能理解你这个问题了。推荐张文彤老师《SPSS 统计分析教程》，链接如下：

http://bbs.pinggu.org/thread-3573256-1-1.html。一般的自由度是 DF=N-K-1。其中，N 是样本量，K 是参数个数，最后那个 1 是常数项。祝好运~



提问：

统计报表结果用脚本导出的疑问

请教各位，在文彤老师的那本基础教程里，关于 CCSS 的案例，

有一个是关于将报表结果用脚本导出的：

```
DEFINE M_Tb01(invar1=!charend('/')).
```

```
CTABLES
/VLABELS VARIABLES=!invar1!concat("Q",!invar1)time
DISPLAY=NONE
/TABLE!invar1[C][ROWPCT.COUNT''F40.1]+!concat("Q",!invar1)[S][M
EAN'感受值'F40.1]
BY time[C]
/SLABELS POSITION=ROW
/CATEGORIES VARIABLES=!invar1 ORDER=A KEY=VALUE
EMPTY=INCLUDE
/CATEGORIES VARIABLES=time ORDER=A KEY=VALUE
EMPTY=EXCLUDE.
!ENDDEFINE.
M_tb01 invar1=a3.

SAVE TRANSLATE OUTFILE='D:\OutTbl1.xls'
/TYPE=XLS
/VERSION=8
/MAP
```

/REPLACE

/FIELDNAMES

/CELLS=VALUES.

但是，我用这个脚本，导出数据是 CCSS 的案例数据本身，

并不是生成的统计表结果

不知道是脚本有问题吗？还是我的方法有问题呢？

解 答 栏

据我个人所知，一般大家学 SPSS 都是窗口操作，很少涉及到楼主用的这种代码的。如果要基于代码操作，一般都不会使用 SPSS 而转向使用 STATA,Eviews,Sas,R 等其它软件了。PS：楼主想做的不外乎就是想导出分析结果，直接用窗口把命令实现后，将相关表格复制到 excel 或者 word 应该就行了，这和代码脚本导出一个道理。祝好运~



提问：

SPSS 回归结果中星号标记的问题

SPSS 回归结果中的星号标记都是需要手动整理的吗？

是不是每次选择不同的置信区间 95%90%99%，分别回归，然后看 sig，最后手动添加？

解 答 栏

不用每次选择不同的置信区间（95%、90%和 99%）去分别回归，直接一次回归就行。回归结果出来后，直接看回归系数对应的 sig/p 值。一般文献中，sig<0.1 时，加 1 颗星；sig<0.05 时，加两颗星；sig<0.01 时，加 3 颗星。祝好运~

3.3

STATA

数据的录入与储存

STATA 为用户提供了简捷，但是非常完善的数据接口，熟悉它的用法是使用 STATA 的第一步，在 STATA 中读入数据可以有三种方式：直接从键盘输入、打开已有数据文件和拷贝、粘贴方式交互数据。

3.3.1 从键盘输入数据

在 STATA 中可以使用命令行方式直接建立数据集，首先使用 input 命令制定相应的变量名称，然后一次录入数据，最后使用 end 语句表明数据录入结束。

例 1 在某实验中得到如下数据，请在 STATA 中建立数据集。

观测数据:X 1 3 5 7 9,Y 2 4 6 8 10

解：此处需要建立两个变量 X、Y，分别录入相应数值，STATA 中的操作如下，其中划线部分为操作者输入部分。

```
. drop  all

. input x y
        x     y
  1. 1     2
  2. 3     4
  3. 5     6
  4. 7     8
  5. 9    10
  6. end
```

3.3.2 用 STATA 的数据编辑工具

①进入数据编辑器

进入 stata 界面，在命令栏键入 edit 或在 stata 的 window 下拉菜单中单击 data editor 或点击编辑图标 (注意：是浏览图标，点击后只能浏览，不能编辑)即可进入 stata 数据编辑器。（stata 界面如下图 2）

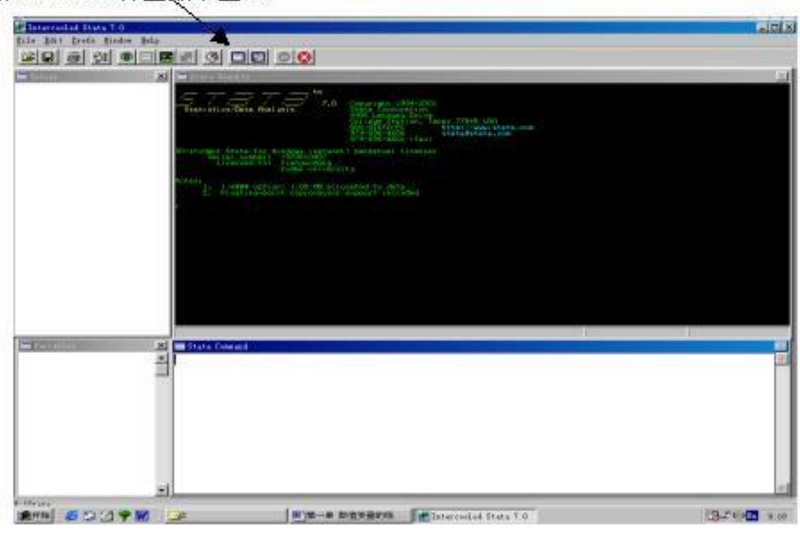


图 2

②数据编辑

STATA 数据编辑器界面：此时进入了数据全屏幕编辑状态。

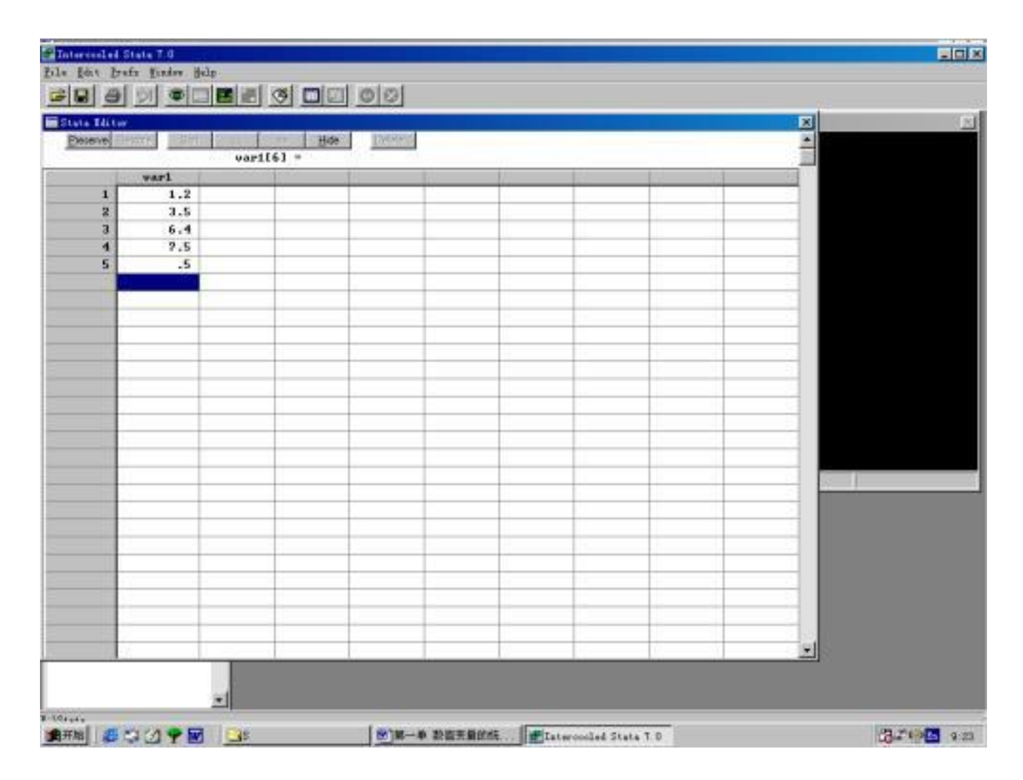


图 3

在第一列输入数据后，STATA 第一列自动命名为 var1；在第二列输入数据后，第二列自动命名为 var2……依次类推。在输入数据后，双击纵格顶端的变量名栏(如：Var1 或 Var2 处)，可以更改变量名，并可以在 label 栏中注释变量名的含义，点击 OK 确认（如图 4 所示）。仍沿用上例，双击观察值所在列顶端的变量名栏，更改变量名为 x，并在 label 栏中注明“7 岁男童身高/cm”。

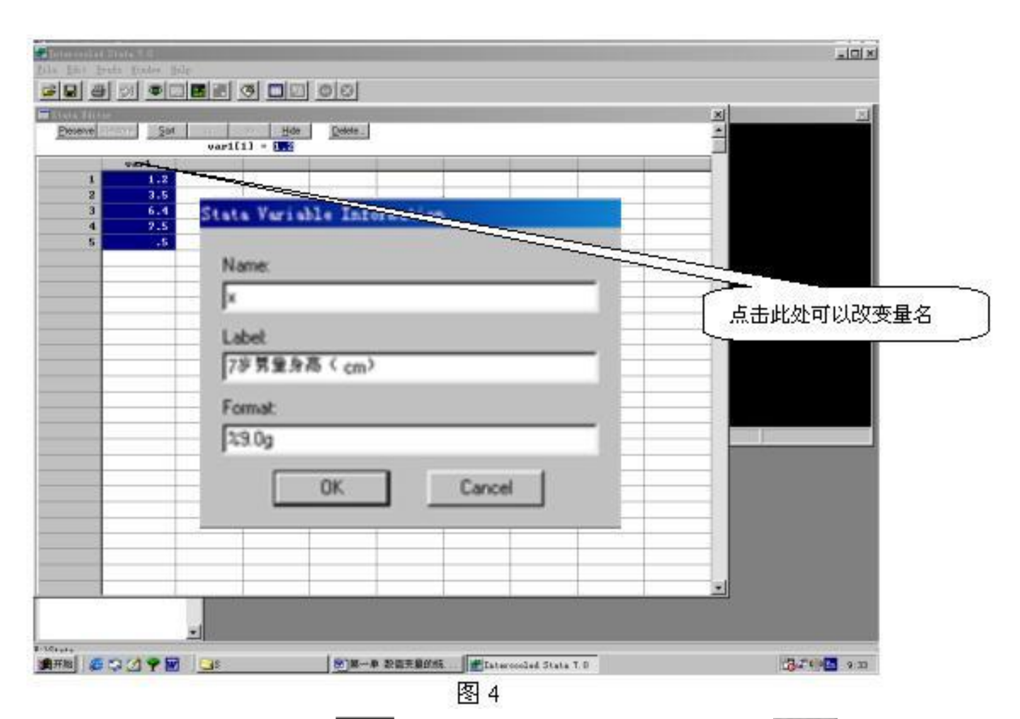


图 4

数据输入完毕后，单击 preserve 键确认所输数据，按关闭键 X 即可退出编辑器。

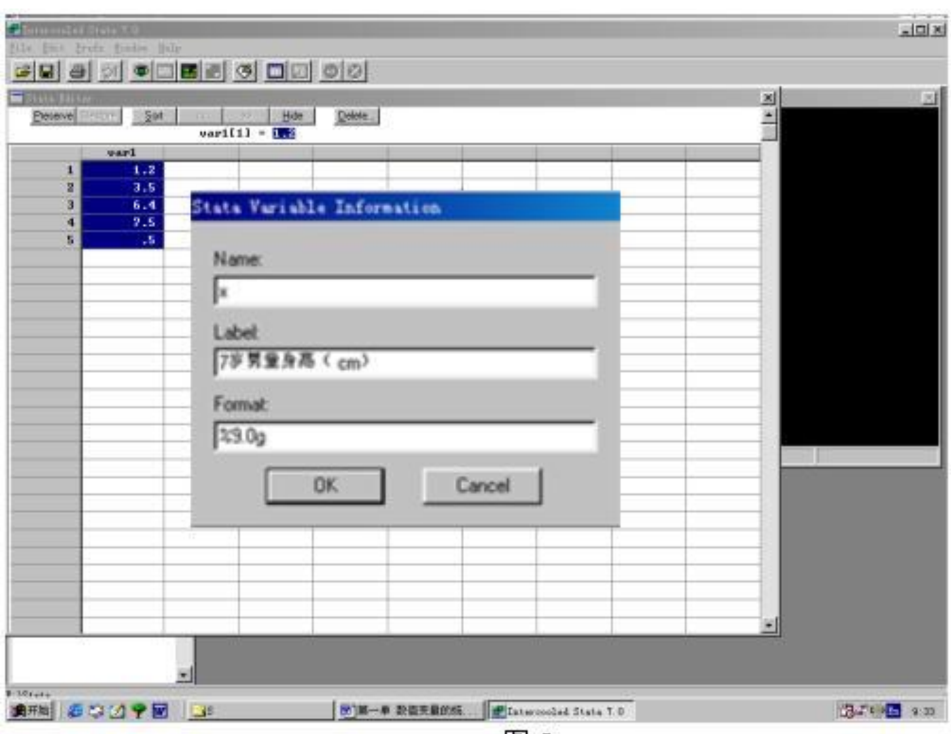


图 5

数据输入完毕后，单击preserve认所输数据，按关闭键×即可退出编辑器。

3.3.3 拷贝、粘贴方式交互数据

STATA 的数据编辑窗口是一个简单的电子表格，可以使用拷贝、粘贴方式直接和 EXCEL 等软件交互数据，在数据量不大时，这种方式操作极为方便。

例 2 现在 EXCEL 中已录入了三个变量，共五条记录，格式见下图，请将数据读入 STATA。

解：首先将 EXCEL 中的 A1 - C6 全部 18 个单元格选中，选择菜单编辑è复制，将数据拷贝到剪贴板上；然后切换到 STATA，选择菜单 Window->Data Editor，打开数据编辑窗口；再选择 Edit->Paste，相应的数据就会被直接粘贴如数据编辑窗口中，并且变量名、记录数、变量格式等均会被自动正确设置，见图 6 和图 7。

	A	B	C
1	x	y	z
2	1	2	q
3	3	4	wqw
4	5	6	e
5	7	8	dfw
6	9	10	f

图 6　在 EXCEL 中的数据格式

	x	y	z
1	1	2	q
2	3	4	wqw
3	5	6	e
4	7	8	dfw
5	9	10	f

图 7　粘贴入 Stata 后的数据格式

3.3.4 打开已有的数据文件

STATA 能够直接打开的数据文件只能是自身专用格式或者以符号分隔的纯文本格式，后者第一行可以是变量名，分述如下：

①点击图标，然后选择路径和文件名，可以打开 STATA 专用格式的数据文件，并且扩展名为.dta。

②打开 Dta 数据文件：该格式文件是 STATA 的专用格式数据文件，也使用 use 命令即可打开，例如要打开数据文件“C:\data1.dta”，则命令为：

.use c:\data1

即扩展名可以省略，如果 STATA 中已经修改或者建立了数据集，则需要使用 clear 选项清除原有数据，命令为：

.use c:\data1,clear

②读入文本格式数据：需要使用 insheet 命令实现，例如需要读入已建立好的文本格式数据文件“C:\data1.txt”，则命令为：

.insheet using c:\data1.txt

该命令会自动识别第一行是否为变量名，以及变量列间的分隔符是 tab、逗号还是其他字符。如果 STATA 中已经修改或者建立了数据集，则需要使用 clear 选项清除原有数据，命令为：

.insheet using c:\data1.txt,clear

3.3.5 数据文件的保存

为了方便以后重复使用，输入 STATA 的数据应存盘。STATA 实际上只能将数据存为自身专用的数据格式或者纯文本格式，分述如下：

①点击图标，然后选择路径和文件名，点击　　。

②存为 dta 格式：可以直接使用文件菜单，也可以使用 save 命令操作，如欲将上面建立的数据文件存入“C:\”中，文件名为 Data1.dta，则命令为：

.save c:\data1

file c:\data1.dta saved

该指令将在 C 盘根目录建立一个名为“data1.dta”的 STATA 数据文件，后缀 dta 可以在命令中省略，会被自动添加。该文件只能在 STATA 中用 use 命令打开。如所指定的文件已经存在，则该命令将给出如下信息：file c:\data1.dta already exists，告诉用户在该目标盘及子目录中已有相同的文件名存在。如欲覆盖已有文件，则加选择项 replace。命令及结果如下：

.save c:\data1.dta,replace

file c:\data1.dta saved

3．存为文本格式：需要使用 outsheet 命令实现，该命令的基本格式如下。

outsheet[变量名列表]using 文件名[,nonames replace]

其中变量名列表如果省略，则将全部变量存入指定文件。如欲将上面建立的数据文件存入文本文件“C:\data1.txt”中，则命令为：

.outsheet using c:\data1.txt

此时建立的文件 data1.txt 第一行为变量名，第 2-6 行为变量值。变量列间

用 Tab 键分隔。如果不希望在第一行存储变量名，则可以使用 nonames 选项。

如果文件已经存在，则需要使用 replace 选项。



提问：

STATA 跑数据的问题

我在写一份定量论文，要跑数据跑模型，用的 CHARLS 数据，由于该数据的数据库是分好几块的，我用了其中几块数据，待我将我所需变量都生成之后，发现跑不了数据，我每打开一个版块的数据它会自动带上 clear 命令，所以，在最后跑模型的时候总是提示某个变量没找到，要怎么解决这一问题？mlogit 的具体格式是怎么样的？

解 答 栏

我们在一篇实证分析文章时，一般用的指标都不会太多。所以针对第一个问题我个人的建议是先将需要构建模型的变量整理出来，然后单独将其导入到 STATA 中，存为一个文件，后面的建模过程都基于这个文件实现；至于第二个问题，即 mlogit 的具体格式问题，推荐陈强老师《高级计量经济学及 STATA 应

用》、卡梅伦《用 STATA 学微观计量经济学》等教材，上面有原理介绍，有代码操作，有案例解读。此外，下面附带一些 STATA 处理 mlogit 的帖子，应该对你有用。祝好运~

相关链接：【STATA】关于 mlogit 回归结果的解释

http://bbs.pinggu.org/thread-2175958-1-1.html

如何在 STATA 中 mlogit 将其多个回归方程横向并列的输在一张表格上？

http://bbs.pinggu.org/thread-732600-1-1.html

关于 mlogit 回归中出现 back up 无法 converge 的问题

http://bbs.pinggu.org/thread-2597055-1-1.html

STATA 命令分类列表大全（日常背诵非常好）

http://bbs.pinggu.org/thread-1528489-1-1.html

STATA 由浅及深全套课件及初级教科书

http://bbs.pinggu.org/thread-3332539-1-1.html

winsor2-批量进行 winsor 或 trimming 处理

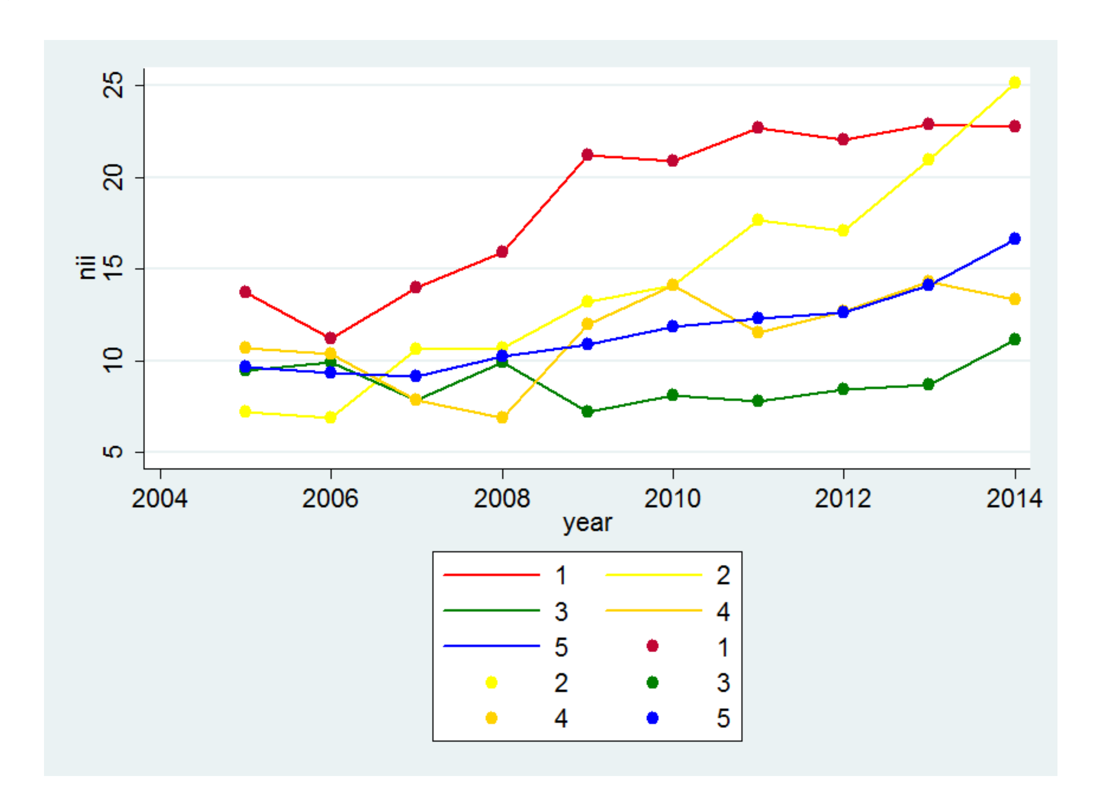
http://bbs.pinggu.org/thread-2807843-1-1.html



提问：

用 STATA 做的图，麻烦大神给指点一下

鄙人刚开始搞论文，以前也没投过稿，现在第一篇。这是我用 STATA 做的图，一共五个样本，做到了一张图上，但是自己感觉乱乱的，而且也不好看。麻烦诸位大神能不能指点一下，多谢多谢



解 答 栏

建议楼主使用上面的图，因为上面的图占篇幅要小一些。不过在投稿前，建议对图进行相应的美化，比如把 lable 加到上面的图列框里去，然后把上面的图的时间刻度调一调。同时，如果期刊需要黑白色的图（楼主也可以节约版面费），就用 STATA 提供的黑白图的那种模板套进去。具体实现过程可参考 STATA 的 options 选项设置。下面给你推荐一些基础资料，祝好运~

相关链接：

STATA 表格输出求助

http://bbs.pinggu.org/thread-3876025-1-1.html

几十个 STATA 作图举例

http://bbs.pinggu.org/thread-221786-1-1.html

请问三次项的作图该怎么进行呢

http://bbs.pinggu.org/thread-3817867-1-1.html

STATA 作图技巧--实例讲解

http://bbs.pinggu.org/thread-1417073-1-1.html

STATA 制图 tips：颜色（名称）的图示

http://bbs.pinggu.org/thread-2181805-1-1.html

DID 作图如何写 STATA 命令

http://bbs.pinggu.org/thread-3196496-1-1.html

[我的 STATA 无法作图]——解决方法

http://bbs.pinggu.org/thread-38889-1-1.html

STATA 作图后如何去掉底色

http://bbs.pinggu.org/thread-1612804-1-1.html

STATA 作图指导（英文原版）

http://bbs.pinggu.org/thread-2988669-1-1.html



提问：

STATA 怎么保存啊？

如题，请问有谁可以告知一下，怎么保存 STATA 的运行结果啊？

解答栏

楼主这个问题问的比较宽泛啊，方法很多。“笨”方法是直接将结果直接复制到 word 或者 excel 里去整理，省事一些的方法是利用 STATA 的官方命令输出结果，如使用 outreg2,esout,tabout,logout 等命令。一些相似的问题可参见下面的参考资料，祝好运~

相关链接：

如何配置 STATA 以便保存输入过的命令

http://bbs.pinggu.org/thread-270131-1-1.html

STATA 数据保存为 sav 数据

http://bbs.pinggu.org/thread-3224878-1-1.html

log 文件可以保存图表吗？

http://bbs.pinggu.org/thread-3482565-1-1.html

怎么保存 STATA 分组回归的残差

http://bbs.pinggu.org/thread-1135366-1-1.html

如何输出 STATA 的图，和保存？

http://bbs.pinggu.org/thread-44330-1-1.html

tabout 的输出文件保存在什么地方

http://bbs.pinggu.org/thread-3240510-1-1.htm



提问：

STATA 描述性统计

STATA 中的描述性统计功能怎么实现呢？另外，我的 6 个自变量里有 4 个是虚拟变量，做描述性分析的时候该怎么办？

解 答 栏

http://bbs.pinggu.org/thread-3793980-1-1.html

看看下面链接中的相关内容，应该能帮到你。祝好运~

相关链接：

-summarize-table-tablestat-描述性统计

http://bbs.pinggu.org/thread-3792920-1-1.html

-outreg2-：回归结果输出、描述统计、频率及交叉列连分析

http://bbs.pinggu.org/thread-3810068-1-1.html

STATA 如何做分组描述性统计

http://bbs.pinggu.org/thread-2177512-1-1.html

教你如何用 STATA 按论文的报告要求输出满意的结果（转载）

http://bbs.pinggu.org/thread-3843080-1-1.html

STATA 描述性统计

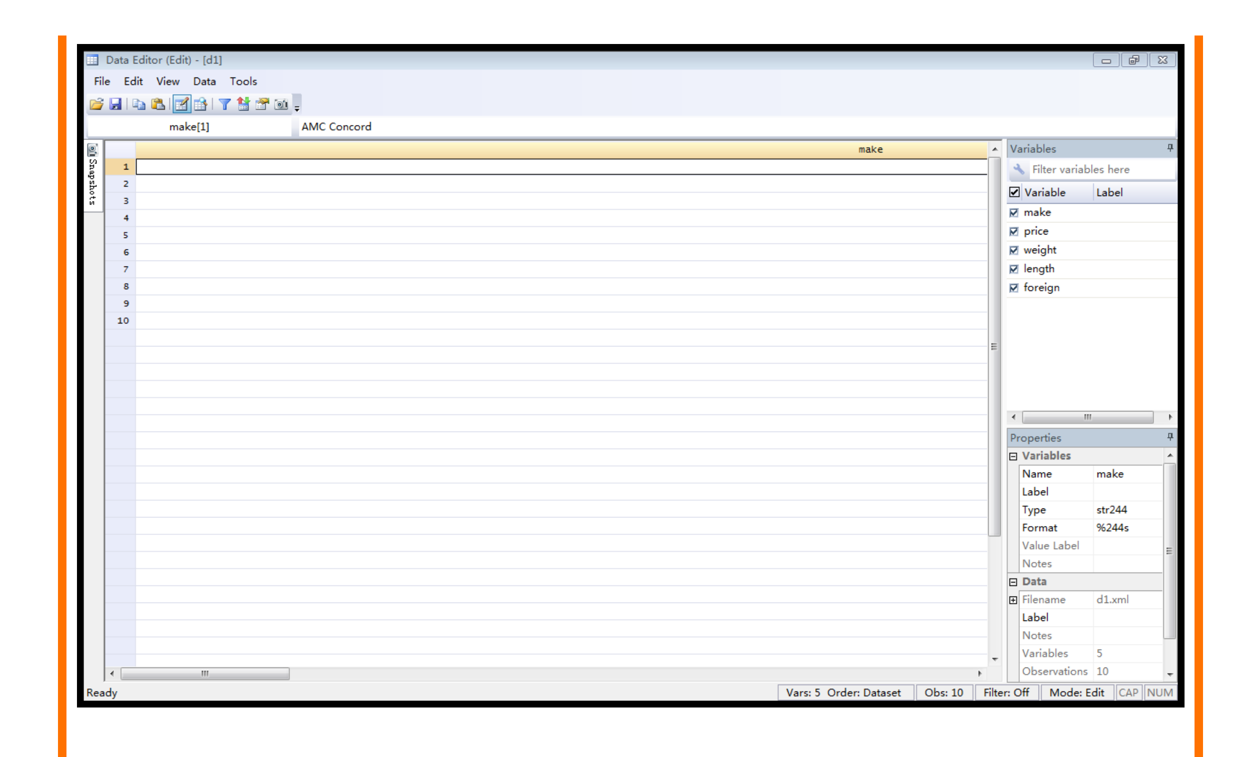
http://bbs.pinggu.org/thread-3793980-1-1.html



提问：

表格在 STATA 里太长了怎么破

我 insheet 了一个文件，但是 browse 时候发现第一个 make 这一列好长，怎么破！谢谢！



解 答 栏

使用 compress 命令即可处理。关于 STATA 的一些常见问题，可参考以下链接。祝好运~

相关链接：

[STATA 数据压缩]你真的会用-compress-吗？

http://bbs.pinggu.org/thread-3566263-1-1.html

初学 STATA 常碰到的问题及解答 100 题

http://bbs.pinggu.org/thread-2921354-1-1.html

STATA 绘图方法大全

http://bbs.pinggu.org/thread-3202543-1-1.html

STATA 的循环

http://bbs.pinggu.org/thread-3202716-1-1.html

初入 STATA 门（浅谈学习方法）转

http://bbs.pinggu.org/thread-2879901-1-1.html



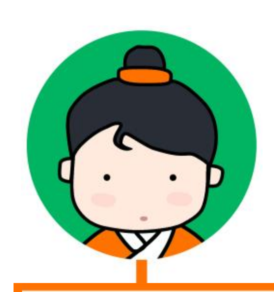
提问：

STATA 中缺省值处理

请教一下各位，STATA 默认的对缺省值是如何处理的？是直接做缺省值，在分析时软件自动删除这个样本吗？

解 答 栏

STATA 做回归等分析时缺失值是直接处理了的。



提问：

STATA 一介差分的标准差怎么做啊

就一个问题，要做一介差分的标准差，但是真心不会这个。

风险（波动性）	RISK	主营业务收入一阶差分的标准差 / 总资产

解 答 栏

要用 STATA 软件求一阶差分标准差。分两边，先做一阶差分；第二步做描述性统计即可，结果会报告标准差的。比如有变量 y，要求其一阶差分：gen y1=L.y；再做 y1 描述性统计分析；des y1 即可。祝好运~



提问：

STATA 双击变量会在 command 里出现变量名，那么 do 文件里怎么才可以快速出现变量名？

STATA 双击变量会在 command 里出现变量名，那么 do 文件里怎么才可以快速出现变量名？在 do 文件里一个个敲变量名工作量太大了。

解 答 栏

无论是 command 窗口还是 do 文件，选择变量进入的过程是必不可少的，是省不了的。a-d 这种写法要求是你 a-d 中间的变量全部进入，但可能回归的时候你选择的变量并不是连着的。另外一种，就是先都在 command 窗口执行完，reviews 窗口就有所以执行过的命令，这时候，全部选中，右键，send to do file 就可以（STATA 的 manual 里面就都介绍过）。祝好运~（注：本问题由原人大经济论坛前蓝色版主回答）



提问：

有没有特别全面的 STATA 教材？麻烦大家推荐

有没有特别全面的 STATA 教材？麻烦大家推荐

解 答 栏

推荐几本我个人看过的，比较好的 STATA 教材，供你参考。陈强《高级计量经济学及 STATA 应用》;卡梅伦《用 STATA 学微观计量经济学》;汉密尔顿著，郭志刚译《应用 STATA 做统计分析》; 廉启国《STATA 数据统计分析教程》等。祝好运~



提问：

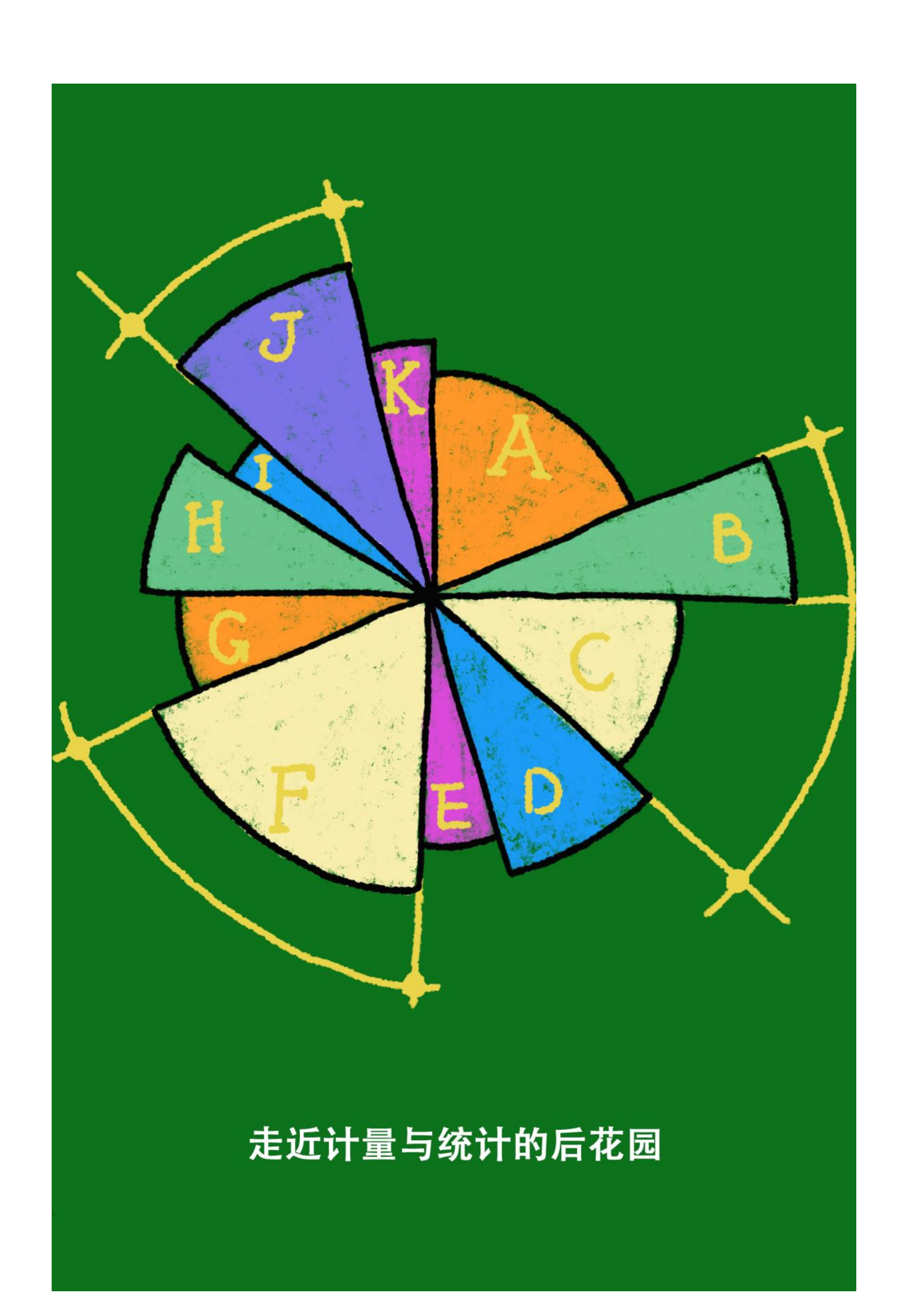
想学习 STATA，但不知从何开始

之前学过 eviews 软件的基本操作，目前想学习一下 STATA，但是不知道从何开始

解 答 栏

介绍点我个人的一点自学 STATA 和计量的经验，仅供参考。首先，STATA 只是一门工具软件，建议在学习它之前先学一些计量的基础原理。推荐伍德里奇《计量经济学导论》,这本书是我看了很多本计量教材后仍然放在桌边的工具书；其次，计量原理用 STATA 实现。如果英文好，建议看 STATA 帮助手册。如果想看中文的教程，统计方面推荐汉密尔顿著，郭志刚翻译的《应用 STATA 做统计分析》, 计量方面推荐陈强《高级计量经济学及 STATA 应用》和卡梅伦著，肖光恩等翻译的《用 STATA 学微观计量经济学》。就我个人来说，觉得以上一本书解决常见的计量/统计问题完全是没问题的；最后，不管是计量的学习还是 STATA

软件的学习，最好的途径还是干中学，边解决实际问题边学习，这样可能更有针对性。祝好运~


J
K
A
I
H
B
G
C
F
E
D
走近计量与统计的后花园

四、

走近计量与统计的后花园：多元分析精彩答疑与学习资源荟萃

〖又遇到一个良心期刊，虽把我拒了，但给了我几页密密麻麻详细的修稿意见，涉及模型的稳健性，变量的内生性问题。那时我才发现以前的心态是多么的可笑，以为看了 SPSS 软件学习培训视频就可以解决以后科研道路上大部分问题，却不知道审稿专家所谓的模型的稳健性和变量的内生性是什么。这让我瞬间有点惶恐，焦灼了一段时间。直到后来各种机缘巧合下慢慢向计量经济学的学习探索之路靠近。〗

4

多元分析

4.1

联合分析&聚类

4.1.1 联合分析

1、定义

市场研究中一个经常遇到的问题是：在研究的产品或服务中，具有哪些特征的产品最能得到消费者的欢迎。一件产品通常拥有许多特征如价格、颜色、款式以及产品的特有功能等，那么在这些特性之中，每个特性对消费者的重要程度如何？在同样的（机会）成本下，产品具有哪些特性最能赢得消费者的满意？要解决这类问题，传统的市场研究方法往往只能作定性研究，而难以作出定量的回答。联合分析（Conjoint Analysis，也译为交互分析）就是针对这些需要而产生的一种市场分析方法。

联合分析法又称多属性组合模型，或状态优先分析，是一种多元的统计分析方法，它产生于 1964 年。虽然最初不是为市场营销研究而设计的，但这种分析法在提出不久就被引入市场营销领域，被用来分析产品的多个特性如何影响消费者购买决策问题。

联合分析是用于评估不同属对消费者的相对重要性，以及不同属性水平给消费者带来的效用的统计分析方法。

联合分析始于消费者对产品或服务（刺激物）的总体偏好判断（渴望程度评分，购买意向，偏好排序等），从消费者对不同属性及其水平组成的产品的总体评价（权衡），可以得到联合分析所需要的信息。

2、基本原理

联合分析是通过假定产品具有某些特征，对现实产品进行模拟，然后让消费者根据自己的喜好对这些虚拟产品进行评价，并采用数理统计方法将这些特性与特征水平的效用分离，从而对每一特征以及特征水平的重要程度作出量化评价的方法。

分析步骤

联合分析通常由以下几部分组成：

1）确定产品特征与特征水平：联合分析首先要对产品或服务的特征进行识别。这些特征与特征水平必须是显著影响消费者购买的因素。一个典型的联合分

析包含 6 - 7 个显著因素。确定了特征之后，还应该确定这些特征恰当的水平，例如 CPU 类型是电脑产品的一个特征，而目前市场上电脑的 CPU 类型主要有：奔腾 II 450，奔腾 II350，赛扬 300 等，这些是 CPU 特征的主要特征水平。特征与特征水平的个数决定了分析过程中要进行估计的参数的个数。

2）产品模拟：联合分析将产品的所有特征与特征水平通盘考虑，并采用正交设计的方法将这些特征与特征水平进行组合，生成一系列虚拟产品。在实际应用中，通常每一种虚拟产品被分别描述在一卡片上。

3）数据收集：请受访者对虚拟产品进行评价，通过打分、排序等方法调查受访者对虚拟产品的喜好、购买的可能性等。

4）计算特征的效用：从收集的信息中分离出消费者对每一特征以及特征水平的偏好值，这些偏好值也就是该特征的“效用”。

5）市场预测：利用效用值来预测消费者将如何在不同产品中进行选择，从而决定应该采取的措施。

4.1.2 聚类分析

聚类分析是将物理或者抽象对象的集合分成相似的对象类的过程。本次实验我将对同一批数据做两种不同的类型的聚类；它们分别是系统聚类和 K-mean 聚类。其中系统聚类的聚类方法也采用 3 种不同方法，来考察对比它们之间的优劣。由于没有样本数据，因此不能根据其数据做判别分析。评价标准主要是观察各聚

类方法的所得到的类组间距离和组内聚类的大小。

分析数据依然采用线性回归所使用的标准化后的能源消费数据。

1、系统聚类——最短距离聚类法（SPSS 实现）

我们在 SPSS 中实现最短距离分析非常简单。单击“分析(A)”-->“分类(F)”-->“系统聚类(H)...”。将弹出如图 1-1 所示的对话框，设置相应的参数即可。



图 1-1 最短距离法

我们的数据已经做过标准化，在“转化值”-->“标准化”选项上选无。

在统计量的聚类成员中选择“无”，因为这是非监督分类，不需要指定最终分

出的类个数。在绘制中选择绘制“树状图”。单击确定，得到以果。

2、K-mean 聚类

K-mean 聚类是用户指定类别数的大样本资料的逐步聚类分析。所谓逐步聚类分析就是先把被聚对象进行初始分类，然后逐步调整，得到最终 K 个分类。K-mean 法对离群点敏感容易扭曲数据分布。

单击“分析(A)”--> “分类(F)”--> “K-均值聚类(K)...”将弹出如图 1-5 所示的对话框，我们根据系统聚类法的经验将 K 选择为 5。迭代次数和系统聚类一样选择 25 次。

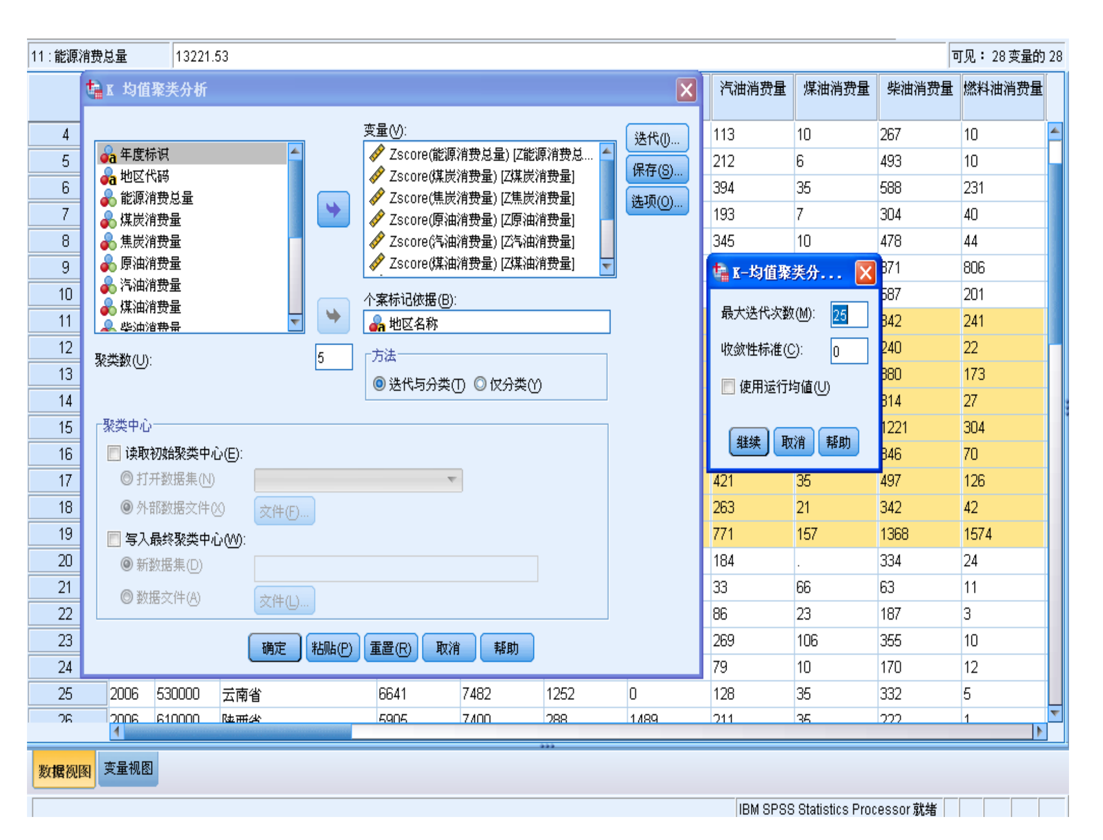


图 1-5 K-mean 聚类设置



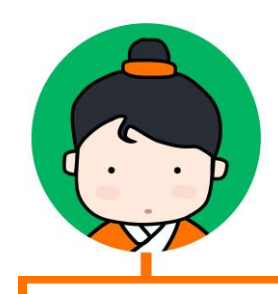
提问：

新手-联合分析问题请教大神们

我最近在做关于网贷方面的投资偏好研究，想用联合分析的方法，已经有数据了，数据内容包括借款人的各种属性：有无照片，借出信用得分，借入信用得分，借款利率，借款时间等等，以及投资人投资的数额，我想用投资人投资数额作为因变量，研究借款人各种属性的重要作用以及组合效用，请问大神们这样可以吗？

解 答 栏

联合分析我没具体实战过，推荐楼主看看这个帖子，看能否有所启发：http://bbs.pinggu.org/thread-3221365-1-1.html。此外，我个人觉得你想做的这个直接用多元线性回归就可实现，要比较各个变量的相对重要性，直接看标准化回归系数即可。祝好运~



提问：

SPSS 聚类分析类间距离如何获得？

用 SPSS16.0 做样品聚类，获得树状图，显示为 0-25 范围，但是我希望获得每类之间的距离，以直观反映类聚合过程，请问如何能获得类与类之间的距离呢？树状图是 text 形式的，是版本问题还是 SPSS 就是这样的呢？急盼复！

解 答 栏

据我个人的了解，类间距离的计算有不同的方法，是相对比较复杂的。所以楼主想弄清楚现在这个问题，我的建议有二：一是具体看你在所用聚类分析时距离的测度勾选的什么方法（一般默认为欧式距离）；二是懂其原理。建议楼主找本介绍聚类分析的教程看看吧，推荐张文彤老师《SPSS 统计分析教程》，链接如下：http://bbs.pinggu.org/thread-3573256-1-1.html。把聚类分析结果输出包含哪些东西看明白了，就明白了其实不用看前面聚类的方法你就知道哪些是一类的，哪些是另一类的了。祝好运~



提问：

SPSS 对取值为 0、1 的不同分类变量做聚类分析？

想请教各位各位 SPSS 高手关于聚类分析的问题，具体的问题是：

对于取值为 0、1 的不同分类变量用哪一种方法做聚类比较合适？如果方便，也麻烦亲提供一下聚类的简单步骤，谢谢啦~

举例：（　衡量是否达标　，达标取 1，否则为 0）

类别 1	类别 2	类别 3 ·
1	0	1
1	1	0
1	1	1

解 答 栏

0,1 这种类别变量可以近似看做连续变量，可以将其以连续变量的形式去做聚类分析。聚类分析的相关知识给楼主推荐两本书，张文彤老师《SPSS 统计分析教程》、《数据分析与挖掘实战案例精粹》上面均有这两种方法的原理介绍和案

例分析，结果解读，论坛有电子版，可找来看看。

http://bbs.pinggu.org/thread-3720908-1-1.html，

http://bbs.pinggu.org/thread-3749834-1-1.html，网址在这里。祝好运~

4.2

主成分&因子

4.2.1 主成分与因子分析的区别与联系

因子分析（Factor analysis）:就是用少数几个因子来描述许多指标或因素之间的联系，以较少几个因子来反映原资料的大部分信息的统计学分析方法。从数学角度来看，主成分分析是一种化繁为简的降维处理技术。

主成分分析（Principal component analysis）:是因子分析的一个特例，是使用最多的因子提取方法。它通过坐标变换手段，将原有的多个相关变量，做线性变化，转换为另外一组不相关的变量。选取前面几个方差最大的主成分，这样达到了因子分析较少变量个数的目的，同时又能与较少的变量反映原有变量的绝大部分的信息。

两者关系：主成分分析（PCA）和因子分析（FA）是两种把变量维数降低以便于描述、理解和分析的方法，而实际上主成分分析可以说是因子分析的一个特例。

两者区别：

（1）因子变量的数量远少于原有的指标变量的数量，因而对因子变量的分析能够减少分析中的工作量。

（2）因子变量不是对原始变量的取舍，而是根据原始变量的信息进行重新组构，它能够反映原有变量大部分的信息。

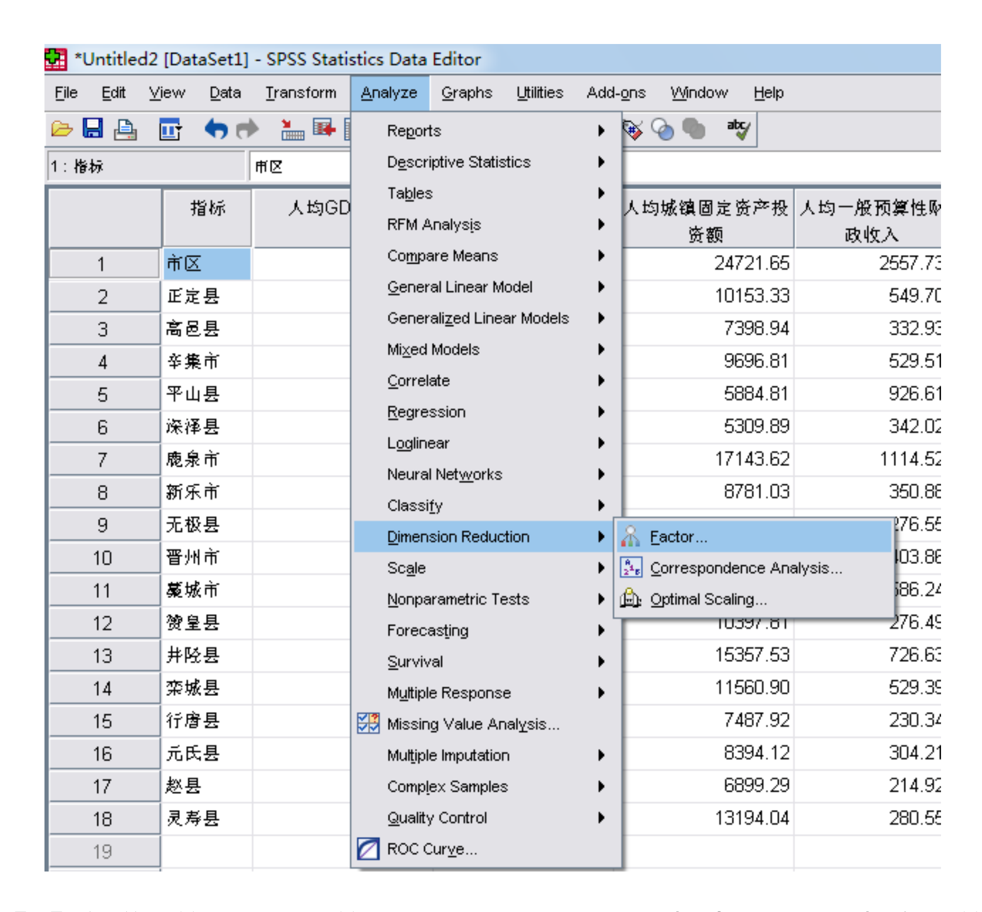
（3）因子变量之间不存在显著的线性相关关系，对变量的分析比较方便，但原始部分变量之间多存在较显著的相关关系。

（4）因子变量具有命名解释性，即该变量是对某些原始变量信息的综合和反映。

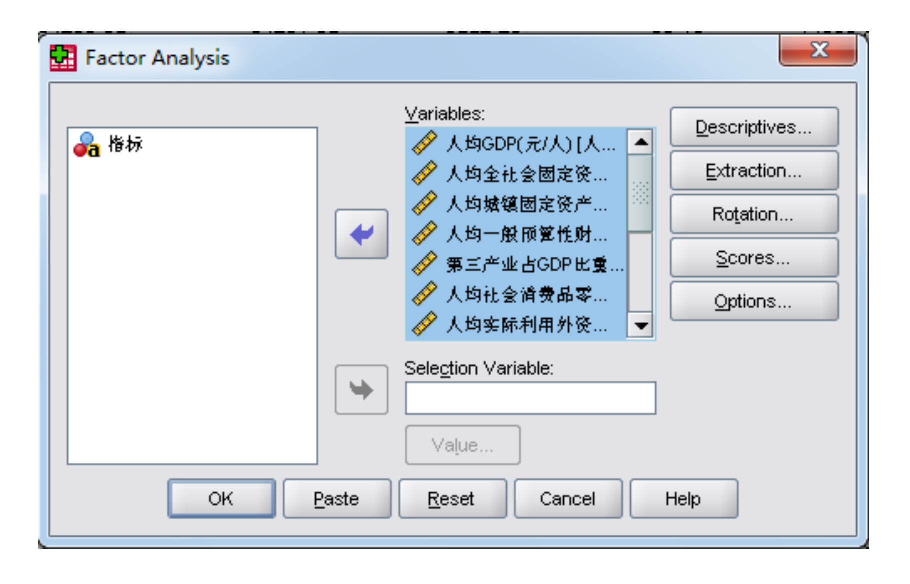
在保证数据信息丢失最少的原则下，对高维变量空间进行降维处理（即通过因子分析或主成分分析）。显然，在一个低维空间解释系统要比在高维系统容易的多。

4.2.2 SPSS 实现

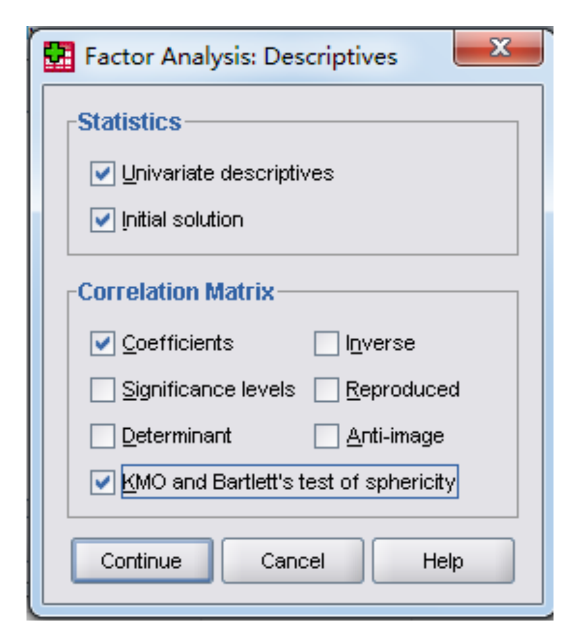
在“Analyze”菜单“Data Reduction”中选择“Factor”命令，如下图所示。



【2】在弹出的下图所示的 Factor Analysis 对话框中，从对话框左侧的变量列表中选择这 14 个变量，使之添加到 Variables 框中。



【3】点击“Descriptives”按钮，弹出“Factor Analysis：Descriptives”对话框，如图所示。



Statistics 框用于选择哪些相关的统计量，其中：

Univariate descriptives（变量描述）: 输出变量均值、标准差；

Initial solution（初始结果）

Correlation Matrix 框中提供了几种检验变量是否适合做引子分析的检验方法，其中：

Coefficients（相关系数矩阵）

Significance leves（显著性水平）

Determinant（相关系数矩阵的行列式）

Inverse（相关系数矩阵的逆矩阵）

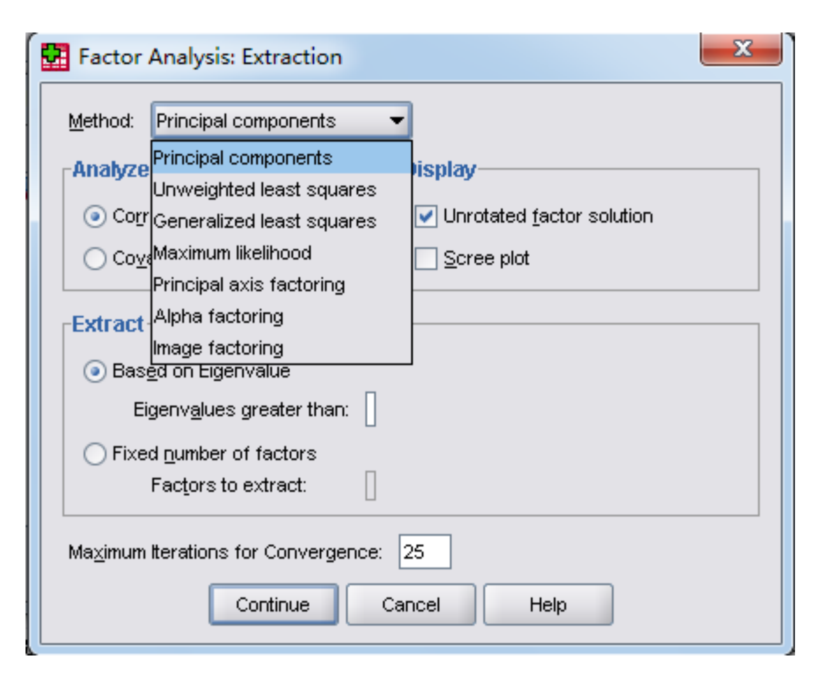
Reproduced（再生相关矩阵，原始相关与再生相关的差值）

Anti-image（反影像相关矩阵检验）

KMO and Bartlett’s test of sphericity（KMO 检验和巴特利特球形检验）

本例中，选中该对话框中所有选项，单击 Continue 按钮返回 Factor Analysis 对话框。

【4】单击“Extraction”按钮，弹出“Factor Analysis：Extraction”对话框，选择因子提取方法，如下图所示：



因子提取方法在 Method 下拉框中选取，SPSS 共提供了 7 种方法：

Principle Components Analysis（主成分分析）

Unweighted least squares（未加权最小平方法）

Generalized least squares（综合最小平方法）

Maximum likelihood（最大似然估价法）

Principal axis factoring（主轴因子法）

Alpha factoring（α因子）

Image factoring（影像因子）

Analyze 框中用于选择提取变量依据，其中：

Correlation matrix（相关系数矩阵）

Covariance matrix（协方差矩阵）

Extract 框用于指定因子个数的标准，其中：

Eigenvaluse over（大于特征值）

Number of factors（因子个数）

Display 框用于选择输出哪些与因子提取有关的信息，其中：

Unrotated factor solution（未经旋转的因子载荷矩阵）

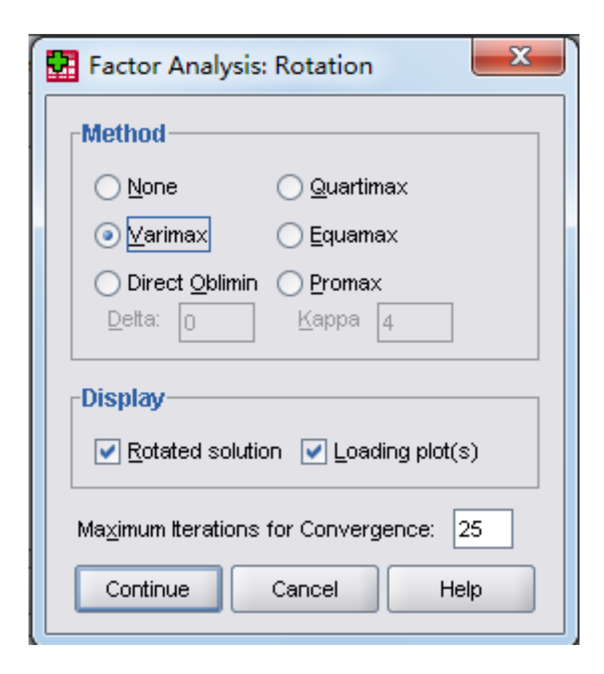
Screen plot（特征值排列图）

Maximun interations for Convergence 框用于指定因子分析收敛的最大迭代次数，系统默认的最大迭代次数为 25。

本例选用 Principal components 方法，选择相关系数矩阵作为提取因子变量的依据，选中 Unrotated factor solution 和 Scree plot 项，输出未经过旋转的因子载荷矩阵与其特征值的碎石图；选择 Eigenvaluse over 项，在该选项后

面可以输入 1，指定提取特征值大于 1 的因子。单击 Continue 按钮返回 Factor Analysis 对话框。

【5】单击 Factor Analysis 对话框中的 Rotation 按钮，弹出 Factor Analysis:Rotation 对话框，如下图所示：



该对话框用于选择因子载荷矩阵的旋转方法。旋转目的是为了简化结构，以帮助我们解释因子。SPSS 默认不进行旋转（None）。

Method 框用于选择因子旋转方法，其中：

None（不旋转）

Varimax（正交旋转）

Direct Oblimin（直接斜交旋转）

Quanlimax（四分最大正交旋转）

Equamax（平均正交旋转）

Promax（斜交旋转）

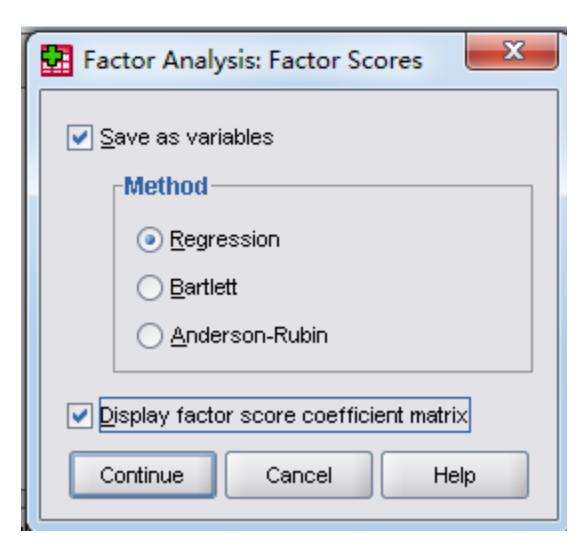
Display 框用于选择输出哪些与因子旋转有关的信息，其中：

Rotated solution（输出旋转后的因子载荷矩阵）

Loading plots（输出载荷散点图）

本例选择方差极大法旋转 Varimax，并选中 Rotated solution 和 Loading plot 项，表示输出旋转后的因子载荷矩阵和载荷散点图，单击 Continue 按钮返回 Factor Analysis 对话框。

【6】单击 Factor Analysis 对话框中的 Scores 按钮，弹出 Factor Analysis:Scores 对话框，如下图所示：



该对话框用以选择对因子得分进行设置，其中：

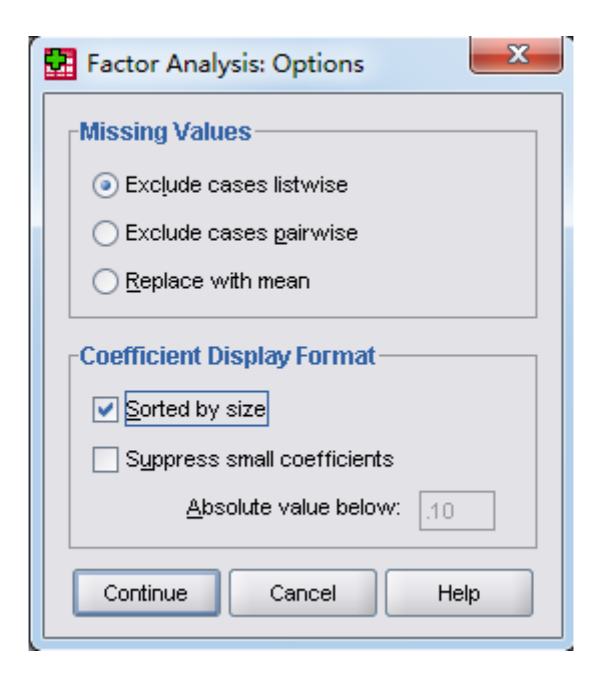
Regression（回归法）：因子得分均值为 0，采用多元相关平方；

Bartlett（巴特利法）: 因子得分均值为 0，采用超出变量范围各因子平方和被最小化；

Anderson-Rubin（安德森-洛宾法）: 因子得分均值为 0，标准差 1，彼此不相关；

Display factor score coefficient matrix：选择此项将在输出窗口中显示因子得分系数矩阵。

【7】单击 Factor Analysis 对话框中的 Options 按钮，弹出 Factor Analysis:Options 对话框，如下图所示：



该对话框可以指定其他因子分析的结果，并选择对缺失数据的处理方法，其中：

Missing Values 框用于选择缺失值处理方法：

Exclude cases listwise：去除所有缺失值的个案

Exclude cases pairwise：含有缺失值的变量，去掉该案例

Replace with mean：用平均值代替缺失值

Cofficient Display Format 框用于选择载荷系数的显示格式：

Sorted by size：载荷系数按照数值大小排列

Suppress absolute values less than：不显示绝对值小于指定值的载荷量

本例选中 Exclude cases listwise 项，单击 Continue 按钮返回 Factor Analysis 对话框，完成设置。单击 OK，完成计算。

4.3

STATA 实现——因子分析

用 2009 年的数据对中国社会发展状况进行综合案例分析。

程序：

```
clear
*定义变量的标签
label var area 省份
label var x1"人均 GDP（元）"
label var x2"新增固定资产（亿元）"
label var x3"城镇居民人均年可支配收入（元）"
label var x4"农村居民家庭人均纯收入（元）"
label var x5"高等学校数（所）"
label var x6"卫生机构数（个）"
describe
```

factor x1-x6

screeplot/*碎石图（特征值等于 1 处的水平线标示保留主成分的分界点）*/

*检验

estat kmo/*KMO 检验，越高越好*/

estat smc/*SMC 检验，值越高越好*/

rotate/*旋转*/

loadingplot,yline(0)xline(0)/*载荷图*/

*预测

predict score fit residual q/*预测变量得分、拟合值和残差以及残差的平方和*/

predict f1 f2

label var f1 收入因子

label var f2"投资、社会因子"

list area f1 f2

summarize f1 f2

correlate f1 f2

scoreplot,xtitle("收入因子")ytitle("投资、社会因子")///

mlabel(area)yline(0)xline(0)/*得分图*/

分析：

首先通过主因子分析(factor)，得到主成分因子：

Factor analysis/correlation Number of obs=31

Method:principal factors Retained factors=3

Rotation:(unrotated)Number of params=15

旋转会进一步简化因子结构。在提取因子之后，键入 rotate 命令进行旋转。

Factor analysis/correlation Number of obs=31

Method:principal factors Retained factors=3

Rotation:orthogonal varimax(Kaiser off)Number of params=15

因子分是通过将每个变量标准化为平均数等于 0 和方差等于 1，然后以因子分系数进行加权合计为每个因子构成的线性组合。基于最近的 rotate 或 factor 结果，predict 会自动进行这些计算。通过命令 predict f1 f2，我们得到了各个观察变量的主因子 1、主因子 2 的得分情况。

.list area f1 f2

.summarize f1 f2

.correlate f1 f2

4.4

STATA 实现——主成分分析

对中国 30 个省市自治区经济发展基本情况的八项指标主成分分析

```
clear*定义变量的标签

label var area 省份

label var x1"GDP(亿元）"

label var x2"居民消费水平(元）"

label var x3"固定资产投资(亿元)"

label var x4"职工平均工资（元）"

label var x5"货物周转量(亿吨公里)"

label var x6"居民消费价格指数(上年 100)"

label var x7"商品零售价格指数(上年 100)"

label var x8"工业总产值(亿元)"

describe
```

pca x1-x8/*主成分估计*/

estat kmo/*KMO 检验，越高越好*/

estat smc/*SMC 检验，值越高越好*/

screeplot/*碎石图（特征值等于 1 处的水平线标示保留主成分的分界点）*/

predict score fit residual q/*预测变量得分、拟合值和残差以及残差的平方和*/

predict f1 f2 f3

predict q1 q2 q3

scoreplot,mlabel(area)yline(0)xline(0)/*得分图 1*/

scoreplot,xtitle("经济社会总量")ytitle("人民生活水平")///

mlabel(area)yline(0)xline(0)/*得分图*/

scatter f2 f3,xtitle("人民生活水平")ytitle("物价水平")///

mlabel(area)yline(0)xline(0)/*得分图*/

scoreplot,factors(3)mlabel(area)/*得分图*/

scoreplot,combined factors(3)mlabel(area)yline(0)xline(0)/*得分图*/

loadingplot,yline(0)xline(0)/*载荷图*/

loadingplot,combined factors(3)yline(0)xline(0)/*载荷图*/

rotate/*旋转*/



提问：

SPSS 做因子分析，最后发现因子综合得分均值为零，什么原因？

SPSS 做因子分析，最后发现因子综合得分均值为零，什么原因？我查阅了很多文献，发现别人的并不为零，请问问题出在哪儿，如何调整，拜求了！请各位大神帮忙答疑。

解 答 栏

据我个人所知，SPSS 软件得到的因子分析综合得分是个标准化的变量，其总体均值为 0，标准差为 1，故而楼主得到这个结果不必奇怪。同时，注意是综合得分总体的均值为 0，单个个体的综合得分有正有负，分别表示大于和小于总体均值。祝好运~



提问：

SPSS 因子分析后的 anova

请问因子分析后用原数据的均值还是因子成分值去做下一步的 Anova 分析啊。

解 答 栏

这个问题可能要看楼主具体想做什么。如果只是想比较原始指标间的差异，直接对原始指标做 Anova 分析即可；如果想比较主成分分析后合成指标间的差异，则用合成的公因子值去做 Anova 分析即可。祝好运~



提问：

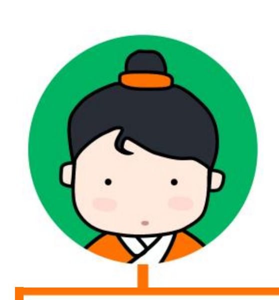
SPSS 怎么用主成分分析后的变量和因变量做一元线性回归

利用主成分分析法，得到由自变量产生的主成分变量，主成分变量是用自变量的一个函数关系式来表示，已知这个变量表达式和因变量，如何建立主成分变量和因变量之间的回归模型？是否利用主成分变量与自变量之间的函数关系，用自变量的数值计算出主成分变量哦数值，这样来建模，这样模型准确不，数据是不是一开始就要归一化，标准化。

解 答 栏

根据我个人对楼主问题的理解，楼主想做的应该是主成分回归，即先利用主成分分析将一些自变量降维，合成一个新的变量，然后再利用这个新的变量去与因变量做回归。所以拆分开来，大概分为以下两步：一是直接对原始变量做主成分分析，得到公因子；二是使用公因子与因变量构建多元线性回归。具体原理介

绍、SPSS 软件实现和案例分析可参考吴明隆老师《问卷统计分析实务——SPSS 操作与应用》第十二章：主成分回归分析与路径分析（412-432），链接如下：http://bbs.pinggu.org/thread-1163508-1-1.html。祝好运~



提问：

有关主成分分析

大家好，我现在一共有 15 个变量，我现在想用主成分分析，选择特征值大于 1 的，结果累积贡献率才 65%，然后我改为提取多个因子，结果提取了 9 个，我感觉这样不太好。有没有办法我可以从那 15 个变量里抽出一部分，使累积贡献率提高，同时提取的因子 4 个左右。

解 答 栏

其一，不知道楼主的数据是问卷式量表还是一般数据（如中宏观统计数据），如果是前者，可能需要做信度检验，然后再做主成分分析；其二，各个专业对主成分累积贡献率的要求不一样，65%在某些专业应该是可以接受的，所以建议楼主看看你这个专业方向文献的结果一般都到多少；其三，SPSS 在进行因子分析时，有专门设置公因子个数的对话框，楼主可以到相应的地方将公因子设置为 4，然后看看结果怎样。PS：我们在做因子分析时，可以不用死盯着特征值大于 1 这个评价标准，0.99 与 1.01 在我看来是没啥差别的。祝好运~



提问：

STATA 主成分分析

目前有 4 个变量，想要提取主成分，在 STATA 中进行主成分分析，分析过程如下：

pca var1 var2 var3 var4，根据特征根选取第一个主成分，故 predict Comp1，生成 comp1，后续用来纳入到回归分析当中。请问操作过程正确么？生成的Comp1 可以直接纳入到回归分析当中么？在论坛里发现有人用score 得分来带入到后续的回归分析之中，其计算方式是 score=proportion1*comp1+proportion2*comp2 ,请问这两个方式有什么区别？非常感谢！

```
. pca  var1 var2 var3 var4

Principal components/correlation                 Number of obs    =      2922
                                                 Number of comp.  =         4
                                                 Trace            =         4
    Rotation: (unrotated = principal)            Rho              =    1.0000

    --------------------------------------------------------------------------
       Component |   Eigenvalue   Difference         Proportion   Cumulative
    -------------+------------------------------------------------------------
           Comp1 |      3.38456      3.05164             0.8461       0.8461
           Comp2 |      .332915      .124427             0.0832       0.9294
           Comp3 |      .208488      .134445             0.0521       0.9815
           Comp4 |     .0740427            .             0.0185       1.0000
    --------------------------------------------------------------------------

Principal components (eigenvectors)

    --------------------------------------------------------------------
        Variable |    Comp1     Comp2     Comp3     Comp4 | Unexplained
    -------------+----------------------------------------+-------------
            var1 |   0.5079    0.3424    0.5574   -0.5605 |           0
            var2 |   0.5090   -0.4142    0.4205    0.6265 |           0
            var3 |  -0.4944    0.5536    0.5268    0.4141 |           0
            var4 |   0.4884    0.6362   -0.4847    0.3491 |           0
    --------------------------------------------------------------------

. predict Comp1
```

解 答 栏

因为信息没给全，所以没法进一步判断楼主做的正确与否。楼主想做的应该是主成分回归。在做主成分分析时，要先通过球形检验判定数据是否适合做主成分分析（如使用命令 estat kmo），适合做才有楼主的后续操作；同时，做主成分分析的目的是消除变量间的多重共线性，对存在严重多重共线性的指标进行降维。所以楼主可以在做主成分分析前先用自变量做一个多元回归，判断哪些自变量间存在多重共线性。然后直接用主成分分析对存在多重共线性的指标进行合并降维即可。至于楼主的最后一个问题：做下一步的回归分析时是用合并后的直接值还是用进一步的合成值，直接用主成分分析得到的值进行下一步回归分析即可。祝好运~



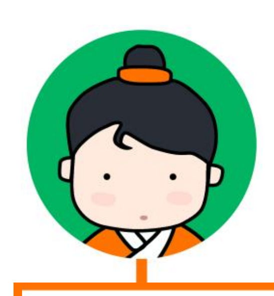
提问：

关于主成分分析中变量的一个问题

请教各位论坛高手，在做主成分分析时，自变量的变化方向的意义不同，也就是说，有的自变量是数值越大对因变量影响越大，即正相关，而有的是数值越小对因变量影响越大，即负相关。这样的变量数据能放在一起做主成分分析么？如果要处理数据，请问怎么处理？

解 答 栏

在我个人看来，自变量对因变量的影响方向不影响其是否能做主成分分析。做主成分分析的目的主要是针对存在严重多重共线性的自变量进行降维，所以建议楼主在做主成分分析前先构建模型对自变量间是否存在严重多重共线性进行诊断。若存在，则进行主成分分析；不存在，就没必要做主成分分析了。同时，SPSS 软件会对变量是否适合做主成分给出判断统计量（如 KMO 值），楼主可以尝试做做，然后看看结果是否适合做主成分分析。祝好运~



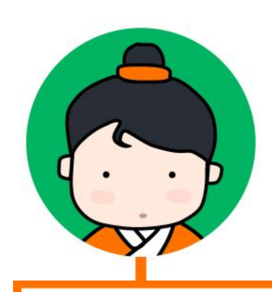
提问：

对提取的主成份进行线性回归的操作

因变量和自变量不对应的可以吗？数据量不对应可以吗？我的因变量是历年的耕地面积，自变量是主成份，我跑出来好像有点不对劲，有哪位知道提取的主成份进行线性回归的具体操作呀。

解 答 栏

根据楼主的描述（历年耕地面积），楼主的数据应该是面板数据。所以即使自变量间存在严重的多重共线性，方法应该使用面板主成分回归。即先对存在严重多重共线性的自变量进行主成分分析，将其合并为公因子，然后以合成的公因子为下阶段模型的自变量，以历年的耕地面积作为因变量，进一步建立回归模型。鉴于是面板数据，后面的回归模型需要在混合 ols 回归、固定效应模型和随机效应模型间进行相关的统计量检验，以确定最后用什么模型。祝好运~



提问：

能不能用因子分析和回归分析做宏观经济分析

比如说我用因子分析来降维归纳解释经济指标然后用回归分析来预测后面短期的一个走势这样做有可行性吗？

解 答 栏

可以，用主成分回归分析方法即可。具体原理介绍、SPSS 软件实现和案例分析可参考吴明隆老师《问卷统计分析实务——SPSS 操作与应用》第十二章：主成分回归分析与路径分析（412-432），链接如下：http://bbs.pinggu.org/thread-1163508-1-1.html。祝好运~



提问：

怎么解释主成分分析的得分图？

怎么解释主成分分析的得分图？

解 答 栏

SPSS 主成分分析可以得到很多图，如碎石图，最终主成分得分分布图，不知道楼主想解释的是哪个图。推荐楼主看看张文彤老师《SPSS 统计分析教程》，上面有对因子分析比较详细的介绍。链接如下：

http://bbs.pinggu.org/thread-3573256-1-1.html，祝好运。


Y
X
a
o
M
¥
I
b
C
I
采撷计量与统计的蜜果

五、

采撷计量与统计的蜜果：回归分析精彩答疑 1 与资源荟萃

〖我是在各种机缘巧合下开启计量经济学的学习探索之路。仔细想来，实际是解决两个现实问题的迫切需要促使我向这条路靠近。一是稿件被拒后审稿意见中的模型的稳健性和变量的内生性看不懂的问题，这两个概念已经超过当时我的认知了；二是我自己买的几本专业方面的专著中广泛的涉及到了 FGLS，2sls 等模型，仅仅学了张老师 SPSS 培训视频和看过几本 SPSS 统计书籍的我对这些概念和软件实现更是不懂，看着费力。这两个问题着实困扰了我一段时间，直到机缘巧合下伍德里奇《计量经济学导论（第三版）》的出现。〗

5.1

变量问题

5.1.1 变量问题

变量一般在原始数据库里已有，通常也在导入软件时已做类型设置。这里我们所讲的变量设置，主要是指根据建模需要所添加的调节变量和虚拟变量。

什么是调节变量？

如果变量 Y 与变量 X 的关系是变量 M 的函数，称 M 为调节变量。就是说，Y 与 X 的关系受到第三个变量 M 的影响。调节变量可以是定性的(如性别、种族、学校类型等)，也可以是定量的(如年龄、受教育年限、刺激次数等)，它影响因变量和自变量之间关系的方向(正或负)和强弱。 例如，学生的学习效果和指导方案的关系。往往受到学生个性的影响：一种指导方案对某类学生很有效，对另一类学生却没有效，从而学生个性是调节变量。又如，学生一般自我概念与某项自我概念(如外貌、体能等)的关系，受到学生对该项自我概念重视程度的影响：很重视外貌的人，长相不好会大大降低其一般自我概念：不重视外貌的人，长相不

好对其一般自我概念影响不大，从而对该项自我概念的重视程度是调节变量。

- **什么虚拟变量？**

虚拟变量又称虚设变量、名义变量或哑变量，用以反映质的属性的一个人工变量，是量化了的自变量，通常取值为 0 或 1。引入哑变量可使线形回归模型变得更复杂，但对问题描述更简明，一个方程能达到两个方程的作用，而且接近现实。例如，反映文化程度的虚拟变量可取为：1：本科学历；0：非本科学历一般地，在虚拟变量的设置中：基础类型、肯定类型取值为 1；比较类型，否定类型取值为 0。

什么是异常值？

异常值顾名思义，是异常大或异常小的观测值。 异常值可能对统计结果（如均值）产生不相称的影响，并因此导致误导性解释。例如，一个数据集包含值：1、2、3、34。均值 10 受到了极值点 34 的很大影响，大于数据中的大多数值 (1、2、3)。在此例中，该均值使得这些数据值显得比实际值要大。您应该调查异常值，因为他们能够提供关于数据或过程的有用信息。出现异常值的原因有多种：

1. 数据输入错误：纠正错误并重新分析数据
2. 过程问题：调查过程找出出现异常值的原因
3. 缺失因子：确定是否没有考虑影响过程的某个因子

随机偶然性：调查过程和异常值，以确定是否偶然出现；

执行包含及不包含异常值的分析以了解它对结果的影响通常，通过图形来识别异常值是最容易的。



提 问：

SPSS 回归分析中如何将自变量按时间分类进入模型？

分析结果和原数据如图，数据中，有 year_2009 -> year 2014 6 个变量来标记日期。如该行数据为 2009 年的数据，则 year_2009 的值为 1，其余 5 个变量的值为 0。求问各位老师大神最后结果是怎么用 SPSS 实现的？

ROA	数值(N)	8	4	ROA (%)	无	无	8
Board_size	数值(N)	8	0	Board Size	无	无	8
Duality	数值(N)	8	0	Duality	{0, non-duality}...	无	8
Board_inde...	数值(N)	8	2	Board Independ...	无	无	8
Managerial_...	数值(N)	8	4	Managerial Ow...	无	无	8
Leverage	数值(N)	8	2	Leverage (debt/...	无	无	8
Firm_size	数值(N)	8	2	Firm Size	无	无	8
Firm_age	数值(N)	8	2	Firm Age	无	无	8
Liqudity	数值(N)	8	2	Liquidity	无	无	8
Year_2009	数值(N)	8	0	2009	{2009, 1}...	无	8
Year_2010	数值(N)	8	0	2010	{2009, 0}...	无	8
Year_2011	数值(N)	8	0	2011	{2009, 0}...	无	8
Year_2012	数值(N)	8	0	2012	{2009, 0}...	无	8
Year_2013	数值(N)	8	0	2013	{2009, 0}...	无	8
Year_2014	数值(N)	8	0	2014	{2009, 0}...	无	8
Manufacture	数值(N)	8	0	Manufacture	无	无	8
Real_Estate	数值(N)	8	0	Real_Estate	无	无	8

解 答 栏

仔细看了一下，楼主给出的表格做的应该是混合 ols 回归，以最后一年为参照组，其它各年以虚拟变量的形式纳入模型。虚拟变量在 SPSS 中的设置可以使用 recode 命令，具体可参见张文彤老师《SPSS 统计分析教程》里面虚拟变量设定相关部分。链接如下 :http://bbs.pinggu.org/thread-3573256-1-1.html。此外，楼主你的数据类型是面板数据，应该在混合 OLS 回归、固定效应模型和随机效应模型间根据检验统计量选择合适的方法。固定效应模型和随机效应模型 SPSS 不好实现，推荐 STATA。书籍推荐陈强老师《高级计量经济学及 STATA 应用》（链接：http://bbs.pinggu.org/thread-3975548-1-1.html ）和肖光恩等翻译的《用 STATA 学微观计量经济学》（链接：http://bbs.pinggu.org/thread-2291053-1-1.html ）。祝好运~



提 问：

模型消除异方差后，如何引入虚拟变量

一个模型本来没有虚拟变量，在这样的情况下，消除了异方差。之后，如果要引入虚拟变量，是直接输入虚拟变量的值，与其他解释变量作回归么？如果这样的话，异方差是不是白消除了？

解 答 栏

楼主为什么要在处理了异方差后再加入虚拟变量呢？为什么不加入虚拟变量后再检验是否存在异方差呢，如果存在，再有针对性的处理呢？同时，楼主又怎么知道加入虚拟变量后就会使得模型存在异方差呢？在我个人看来，这些都是没有依据的啊！所以，建议楼主的先考虑好模型怎么构建，即模型到底需要哪些自变量，然后再构建模型，然后再考虑模型的异方差，内生性问题。

追 问：

这个是我们老师要求我们建立模型，先消除异方差，再引入虚拟变量，具体

也不知道为何。

回　答：

楼主老师比较奇怪。如果是这样的话，建议楼主考虑以下几点：一是知道你所用的软件如何将类别变量处理为虚拟变量；二是加入虚拟变量后的模型再次检验异方差的存在；三是如果存在异方差，要怎么处理；四是关于虚拟变量的解释。祝好运~



提 问：

遗漏变量问题！！

在做回归分析时如果遗漏变量会产生自相关，比如说我现在想研究创新与竞争力之间的关系，可是影响竞争力的又不是只有创新。那我现在是不是不可以做回归？如果看创新对竞争力的影响系数该用什么方法？

解答栏

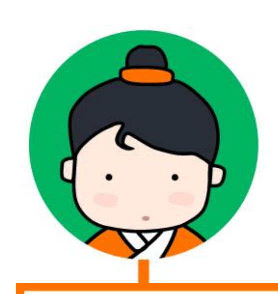
遗漏变量的问题很常见。其实，根据常识，我们基本上是没办法穷尽影响每个因变量的自变量的。然而，楼主你这个还好，你只是想探究创新对竞争力的影响，虽然得到的结果会有一些偏误，但如果根据国内外文献加了一些控制变量后，这些偏误会变小的。故而，可以构建模型，即 Y(竞争力)=回归系数*创新+回归系数*控制变量（这里控制变量可以有多个）+u(扰动项)。大概看看创新对竞争有何影响。

追　问

之前看的文章比如说创新对某一产业的竞争力的影响，是用面板数据直接做的回归，没有任何控制变量，而且我觉得因为不是企业，只是想看对某一产业的影响，这样是不是没有办法添加控制变量?既然没办法添加控制变量的话，请问这样是要看 R 么？问题比较小白~求大神帮助~

回　答

即使是用面板数据做，应该也有控制变量。不至于一个模型就只有创新一个自变量的，应该还有其它的变量，而其它的变量就是所谓的控制变量。根据常识，创新对一个产业的影响，应该不止影响一年。所以用面板数据处理还是比较好的。



提 问：

哑变量结果解读

想请教一个问题：

在 SPSS 中采用多元线性回归分析，自变量有连续变量和无序等级变量，将职业等无序多分类变量经过哑元化处理，将同一因素下的哑变量进行归组，在纳入方法中选择了“ENTER”能够确保这些哑变量同进同出吗？另外结果怎么解释，谢谢。

解 答 栏

类别变量处理为哑变量后，使用 enter 法能保证变量的同进同出的。至于结果的解释，类别变量会有一个参照组，一般都是 xx 组相对于参照组 xx 怎样变化，是参照组的 xx 倍。解释这块建议你看看你这个专业的相关文献，看看别人是怎么解释的，仿着做就行。祝好运~



提 问：

同时存在两个虚拟变量的模型，哪个为默认的基准组？

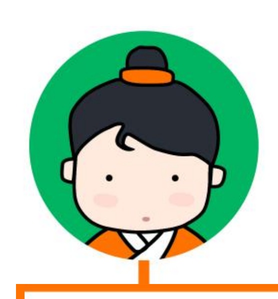
第一个方程包含截距项与 1 个虚拟变量，第二个方程不含截距项但有 2 个虚拟变量，对比两个方程参数的估计结果与标准误，它们有什么相同或相异之处？两个方程的基准组分别是什么？

$$wage = \beta_1 female + \beta_2 educ + \beta_3 exper + \beta_4 tenure + \beta_0 + \varepsilon$$

$$wage = \beta_1 female + \beta_2 educ + \beta_3 exper + \beta_4 tenure + \beta_5 male + \varepsilon$$

解 答 栏

仔细的对比了一下楼主的两个方程，“性别”变量应该是模型关注的变量。具体的对比看楼主怎么对“性别”进行编码的。由于“性别”是 0,1 型的二分类变量，所以建议楼主将“性别”编码问题理清楚后，将具体的编码（0,1）带入两个方程，从截距和斜率两个角度去对比下方程差异就可以了。祝好运~



提 问：

混合线性模型的因子能不能是连续变量？

看有的文献自变量针对连续变量分析，而且还有 95%的 CI，不知道是怎么得出的？连续变量做 factors ,如果选取这个连续变量作为固因子或主效应的话，最后得出的结论是各值的比较，应该怎么理解呢？

解 答 栏

建议楼主看张雷老师的《多层线性模型应用》或者郭志刚老师翻译的《分层线性模型：应用与数据分析方法》，论坛上都有电子版的。里面有关于这个的原理和案例介绍(自然也包括连续性自变量的解释)。95%的 CI(置信区间)在 SPSS 回归的子对话框中可以勾选的。连续变量做 factors？没有懂楼主想表达的意思。混合线性模型不管因变量是连续变量还是类别变量都可以做的。只是当因变量为连续变量时，用的是多元回归；因变量是类别变量时，用的是 logistic 回归。

追　问：

想问自变量是连续变量的时候 95%的 CI 怎么取，俺是初学者，很是懵懂啊！！得出都是与冗余的比较值。。。。。。

回　答：

SPSS“分析”菜单下，选择“回归”，“回归”下选择“线性”，“线性”下选择“保存”，这个子对话框里有“预测区间”，这里可以选择置信区间。



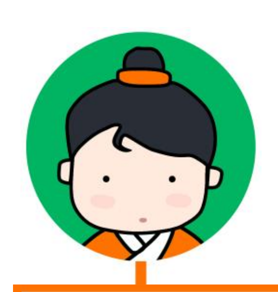
提 问：

求问找对照组和不找对照组有什么区别

看了很多财会类文献，为了验证某一处理的效果，会按照同行业同规模找对照样本。然后设置虚拟变量 D，接受处理的 D=1，不接受处理的 D=0。我想问一下，如果不找对照组，将所有公司混合，接受处理 D=1，不接受处理 D=0，虽然 D=0 这一组会多很多样本，但到底能不能这样做呢？？我是菜鸟，可能涉及到计量的东西确实不太懂，麻烦各位给说说。。。

解 答 栏

就我个人所知道的而言，楼主这个问题应该不简单的是能否将某个“处理”设置为虚拟变量的问题，它应该涉及到了一种相对比较高级的计量方法——倾向匹配值得分。所以建议楼主去查查倾向匹配得分方法的资料，把那个方法的原理搞懂了再来用软件实现。推荐陈强老师《高级计量经济学及 STATA 应用》(第二版)。祝好运~



提 问：

关于异常值的处理

我想做一个医学指标的诊断标准，样本量 10000 多，发现这一指标的影响因素，譬如身高、体重有 575 个异常值，5%的异常率，异常值如何处理才能保证诊断标准比较科学？

解 答 栏

由于你样本量比较大，所以如果是做回归，可以考虑用分位数回归去探究最后 5%群体的特征的。winor 这个命令会使得样本信息损失很多的。

追　问：

是专门把异常值提出来吗？回归分析哪个是分位数回归呢？

回　答：

如果你专业上认为极端异常值分析是有意义的，就应该将它们也包含进来，一起分析。至于分位数回归的介绍，可看陈强老师《高级计量经济学及 STATA 应用》（第二版）第 26 章（509-517），有原理和案例介绍。祝好运~

5.1.2 交互、中间、调节效应

1、调节变量的定义

变量 Y 与变量 X 的关系受到第三个变量 M 的影响,就称 M 为调节变量。调节变量可以是定性的，也可以是定量的。在做调节效应分析时,通常要将自变量和调节变量做中心化变换。

简要模型：Y = aX + bM + cXM + e

Y 与 X 的关系由回归系数 a + cM 来刻画,它是 M 的线性函数, c 衡量了调节效应(moderating effect)的大小。如果 c 显著，说明 M 的调节效应显著。

2、调节效应的分析方法

显变量的调节效应分析方法，分为四种情况讨论。

（1）当自变量是类别变量，调节变量也是类别变量时，用两因素交互效应的方差分析，交互效应即调节效应；

（2）当调节变量是连续变量时，自变量使用伪变量时,将自变量和调节变量中心化,做 Y=aX+bM+cXM+e 的层次回归分析

第一步 做 Y 对 X 和 M 的回归,得测定系数 R12。

第二步 做 Y 对 X、M 和 XM 的回归得 R22，若 R22 显著高于 R12，则调节效应显著。或者,作 XM 的回归系数检验,若显著,则调节效应显著；

（3）当自变量是连续变量，调节变量是类别变量时，分组回归:

按 M 的取值分组,做 Y 对 X 的回归。若回归系数的差异显著,则调节效应显著，

（4）当自变量是连续变量，调节变量是连续变量时，同上做 Y=aX +bM +cXM +e 的层次回归分析。

潜变量的调节效应分析方法，分两种情形：

（1）调节变量是类别变量,自变量是潜变量

当调节变量是类别变量时,做分组结构方程分析。做法是,先将两组的结构方程回归系数限制为相等,得到一个χ^2值和相应的自由度。然后去掉这个限制,重新估计模型,又得到一个χ^2值和相应的自由度。前面的χ^2减去后面的χ^2得到一个新的χ^2,其自由度就是两个模型的自由度之差。如果χ^2检验结果是统计显著的,则调节效应显著；

（2）调节变量和自变量都是潜变量

当调节变量和自变量都是潜变量时，有许多不同的分析方法，最方便的是 Marsh,Wen 和 Hau 提出的无约束的模型。

3、中介变量的定义

自变量 X 对因变量 Y 的影响,如果 X 通过影响变量 M 来影响 Y,则称 M 为中介变量。Y=cX+e1, M=aX+ e2 , Y= c′X+bM+e3。其中,c 是 X 对 Y 的总效应,ab 是经过中介变量 M 的中介效应,c′是直接效应。当只有一个中介变量时,效应之间

有 c=c′+ab，中介效应的大小用 c-c′=ab 来衡量。

4、中介效应分析方法

中介效应是间接效应,无论变量是否涉及潜变量,都可以用结构方程模型分析中介效应。

步骤为：第一步检验系统 c，如果 c 不显著，Y 与 X 相关不显著，停止中介效应分析，如果显著进行第二步；

第二步一次检验 a，b，如果都显著，那么检验 c′，c′显著中介效应显著，c′不显著则完全中介效应显著；如果 a，b 至少有一个不显著，做 Sobel 检验，显著则中介效应显著，不显著则中介效应不显著。

Sobel 检验的统计量是 z=^a^b/sab ，中 ^a, ^b 分别是 a, b 的估计, sab=^a2sb2 +b2sa2, sa,sb 分别是 ^a, ^b 的标准误。

5. 调节变量与中介变量的比较

	调节变量 M	**中介变量 M**
研究目的	X 何时影响 Y 或何时影响较大	X 如何影响 Y
关联概念	调节效应、交互效应	中介效应、间接效应
什么情况下考虑	X 对 Y 的影响时强时弱	X 对 Y 的影响较强且稳定
典型模型	Y=aM+bM+cXM+e	$M=aX+e_2$ $Y=c'X+bM+e_3$
模型中 M 的位置	X,M 在 Y 前面,M 可以在 X 前	M 在 X 之后、Y 之前

	面	
M 的功能	影响 Y 和 X 之间关系的方向(正或负)和强弱	代表一种机制,X 通过它影响 Y
M 与 X、Y 的关系	M 与 X、Y 的相关可以显著或不显著(后者较理想)	M 与 X、Y 的相关都显著
效应	回归系数 c	回归系数乘积 ab
效应估计	^c	^a^b
效应检验	c 是否等于零	ab 是否等于零
检验策略	做层次回归分析,检验偏回归系数 c 的显著性(t 检验);或者检验测定系数的变化(F 检验)	做依次检验,必要时做 Sobel 检验



提 问：

SPSS 线性回归的调节变量？

我在看一篇文章，文章中用线性回归验证调节变量。第一个模型中加入了变量 A。第二个模型中加入了变量 A 和调节变量 B。第三个模型中加入了变量 A、调节变量 B 以及 A、B 的乘积项。这篇文章的回归结果中，这三个调整 R 方的变化趋势有的是依次递减、有的是先减后增、有的是依次递增。一般来讲，调整 R 方不是随着变量个数增多而减小吗？这篇文章这样的结果是否值得怀疑？这三个模型的调整 R 方在理论上有什么关系呢？

解 答 栏

不是调整 R^2 随着加入变量个数增加而增加，是 R^2，调整 R^2 的变化是加入新变量后其对模型解释力的影响变化。推荐你看看邱皓政老师《量化研究与统计分析——SPSS 数据分析范例解析》一书第十二章，该章是专门讲调节效

应的，还有 SPSS 操作截图和结果解释。对帮助你理解这个问题很有帮助。同时，推荐谢宇老师《回归分析》一书交互项这一章节介绍（链接：http://bbs.pinggu.org/thread-3065280-1-1.html），堪称经典。祝好运~



提 问：

关于倒 U 型的调节效应

倒 U 型的调节作用，在 SPSS 里面该如何实现？A 与 B 之间是正向关系，想验证 C 对 A、B 间的关系有调节作用，而且是倒 U 型的调节作用。谢谢谢谢~~

解 答 栏

理论上，交互项正在回归模型中可以被视作高次项，其次数等于参与构造交互项的自变量个数。然而楼主现在想引入 C 对 A 与 B 的倒 U 型调节作用，这个做倒是能做，直接以 C 的一次方和平方项分别与 A 和 B 构建交互项即可。然而，在多数情况下，回归模型中三次以上高次项的存在会使我们难以对回归模型的结果做出合理的解释。建议楼主慎用。祝好运~



提 问：

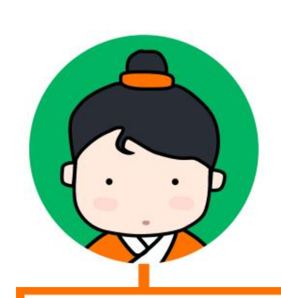
交叉项中存在对数形式，如何解释？

我最近在做污染避难所假说的实证论文。现在遇到一个问题。现用下面一个非常简化的形式来说明我的问题 :logFDI=aB+blogC +c(B*logC), 其中 B 为腐败指数，C 为二氧化碳排放量。考虑到腐败对环境保护是有影响的，因此引入交叉项。可是 C 取了对数，那么若要检测环境变量对 FDI 流入的影响大小，可不可以直接用 logC 前面的系数 b 代入式子：b+c*(B 均值) 来表示二氧化碳排放量对 FDI 的影响？FDI 和 C 都取了对数，应该用弹性解释是吗？谢谢！

解 答 栏

“可不可以直接用 logC 前面的系数 b 代入式子：b+c*(B 均值) 来表示二氧化碳排放量对 FDI 的影响“，我个人觉得可以。正如楼主所说，解释的时候由于楼主关注的自变量和因变量都取了对数，所以用弹性的概念解释即可。关于弹性概念的解释楼主可在论坛找古扎拉蒂《计量经济学基础》（链接：http://bbs.pinggu.org/thread-2924849-1-1.html）或者伍德里奇《计量经济

学导论》（链接：http://bbs.pinggu.org/thread-1152364-1-1.html）看看，里面有案例解释的。祝好运~



提 问：

调节变量是连续变量，自变量是分类变量，交互效应需要中心化吗？

在我的模型中，自变量是 X，因变量是 Y，假设是调节变量 Z 会对 X 对 Y 的效应产生作用，方程是否应该写成： Y=a0+b1*X+b2*X*Z+e

1）对于交互项需要中心化、标准化吗？

2）如果 b1 为正，b2 为负，是否说明 X 的净效应是正向，而调节变量对于 X 对 Y 的作用有削弱作用？

3）看到有帖子说调节变量和交互变量是不一样的，交互是指对称的，调节则是不对称的，因此处理方式有不同，我一直认为这两是一回事，希望有高手帮忙解答下。

解 答 栏

第一个问题，为了避免严重多重共线性的影响，故而在构建交互项前，需要对原始变量进行中心化处理。推荐谢宇老师《回归分析》一书，该书第 13 章是专门讨论交互项的，有关于交互项的中心化（几种不同的中心化方式对结果的影响）。链接：http://bbs.pinggu.org/thread-3065280-1-1.html；

第二个问题，我个人认为不能简单的单独只看 b1 和 b2 的大小和正负，要结合起来一起看，你想做的是看 Z 是否会对 X 产生影响，进而影响 Y，所以对 Y=a0+b1*X+b2*X*Z+e 这个模型取 X 的偏导，得到 b1+b2*Z，要结合这个式子来具体讨论下一步的分析。谢宇老师书里也有相关介绍。

第三个问题，在我个人看来调节变量和交互变量差不多，都是像你构建的这个模型式子一样，只是不同的称呼罢了。祝好运~



提 问：

我想问调节效应是正的，但是调节变量是负的，可以吗？

本来享乐价值和自我效能感对因变量都是正的。现在调节作用中，他俩都负了。虽然达到了正向调节，但是这样有问题吗？

Coefficients[a]

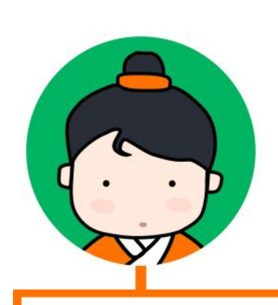
Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	1.180	.175		6.736	.000
	享乐价值	.391	.041	.459	9.462	.000
	自我效能感	.230	.041	.275	5.661	.000
2	(Constant)	2.168	.345		6.294	.000
	享乐价值	-.019	.130	-.022	-.147	.883
	自我效能感	-.104	.109	-.124	-.960	.338
	享乐价值*自我效能感	.140	.042	.626	3.314	.001

a. Dependent Variable: 购买意愿

解 答 栏

楼主在构建模型时，可能未对原始变量先进行中心化处理，然后再构建交互项，这可能使得原始变量和交互项间存在多重共线性，并进而使得加入交互项后

原始变量的偏回归系数方向改变，即出现现在这样的结果(模型 1 中自变量显著，模型 2 中构建交互项后原始变量反而不显著了)。建议先对原始变量去中心化，然后再引入交互项构建模型。关于交互项的讨论，推荐谢宇老师《回归分析》一书，堪称经典。链接如下：http://bbs.pinggu.org/thread-3065280-1-1.html。祝好运~



提 问：

如何解释交互作用？

logistic 估计中，因变量为：是否健康，自变量有：是否富有，城乡等等；估计结果，(是否富有)对(健康)是积极影响。现在我加入了(城乡*是否富有)这个交互变量到上述模型中，估计结果，(是否富有)对(健康)影响变为消极影响(标准化值<1)，(城乡*是否富有)交互变量估计系数的标准化值>1。请问，加入交互变量之后对 (是否富有)变量估计系数的变化说明了什么？

解 答 栏

有交互项，你将类别变量的值分别带入模型，然后看系数对结果进行解释就行。比如你的变量原始编码为 1=城市，0=农村；1=富有，0=不富有。以 0，0（即农村不富有）为参照组，那么分别解释 0,1；1,0；1，1 组结果就行。祝好运~



提 问：

交互作用

因变量为二分类变量

自变量有分类变量，有连续变量

变量间的交互作用怎么做。有交互作用的变量怎么筛选。

解 答 栏

首先，因变量是二分类变量，故而模型选择用二分类 Logistic 回归模型/二分类 probit 模型。

其次，自变量有分类变量，连续变量。分类变量如果是二分类直接纳入模型即可；分类变量是多分类变量，则需转换成虚拟变量/哑变量，然后再纳入模型；连续变量做描述性统计分析，若无极端异常值，基本符合正态分布，直接纳入模型即可。若存在极端异常值，需处理。

最后，构建交互项。利用 compute 功能产生交互项。如你想产生 a 与 b 两个变量的交互项。为避免多重共线性，需先对 a，b 进行中心化处理后，然后再利用 compute 功能产生中心化后的变量的交互项。

虚拟变量的产生过程推荐张文彤老师《SPSS 统计分析教程》，链接：http://bbs.pinggu.org/thread-3573256-1-1.html；交互项的相关知识推荐谢宇老师《回归分析》，链接：http://bbs.pinggu.org/thread-3065280-1-1.html。祝好运~



提 问：

虚拟变量不同基组交叉项系数的显著性

假设有小麦、玉米、稻谷、豆类四类农产品虚拟变量，想看农产品在价格之上的差异。做农产品与价格的交叉项，交乘项系数代表差异，倘若以小麦为基准组，小麦与玉米交叉项显著，但系数较大。同时以玉米为基准组，玉米与小麦的交叉项不显著，说明了什么？显著的交叉项系数较大又是因为什么？

解 答 栏

有可能是构建交互项的原始变量与交互项间存在严重多重共线性导致出现这种结果。建议先对原始变量做对中处理，然后再构建模型，消除多重共线性的影响。祝好运~

5.1.3 OLS（最小二乘法）回归、逐步回归

1、回归分析方法介绍

回归分析是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法，它是我们使用最频繁也是最基础的统计分析方法。

回归分析按照涉及的因变量的多少，分为回归和多重回归分析；按照自变量的多少，可分为一元回归分析和多元回归分析；按照自变量和因变量之间的关系类型，可分为线性回归分析和非线性回归分析。如果在回归分析中，只包括一个自变量和一个因变量，且二者的关系可用一条直线近似表示，这种回归分析称为一元线性回归分析。如果回归分析中包括两个或两个以上的自变量，且因变量和自变量之间是线性关系，则称为多重线性回归分析。

在回归分析中，把变量分为两类。一类是因变量，它们通常是实际问题中所关心的一类指标，通常用 Y 表示；而影响因变量取值的的另一类变量称为自变量，用 X 来表示。

- 回归分析研究的主要问题是：

（1）确定 Y 与 X 间的定量关系表达式，这种表达式称为回归方程；

（2）对求得的回归方程的可信度进行检验；

（3）判断自变量 X 对因变量 Y 有无影响；

（4）利用所求得的回归方程进行预测和控制。

应用回归预测法时应首先确定变量之间是否存在相关关系。如果变量之间不存在相关关系，对这些变量应用回归预测法就会得出错误的结果。

正确应用回归分析预测时应注意：

（1）用定性分析判断现象之间的依存关系；

（2）避免回归预测的任意外推；

（3）应用合适的数据资料。

小福利：回归分析经典教材引荐——

回归分析（谢宇著）

http://bbs.pinggu.org/thread-3065280-1-1.html

《应用回归分析》 何晓群等编著 中国人民大学出版社 2011 年 9 月第 3 版 PDF 教材

http://bbs.pinggu.org/thread-3482911-1-1.html

漫画统计学之回归分析+[日]高桥信

http://bbs.pinggu.org/thread-3832592-1-1.html

线性回归分析基础

《线性回归分析基础》由四种讨论定量方法的小册子组成，分别是《理解回归假设》、《回归诊断简介》、《虚拟变量回归》以及《多元回归中的交互作用》。

http://bbs.pinggu.org/thread-2981212-1-1.html



提 问：

判断 Enter 方式和 Forward Stepwise 方式两个模型哪个

项目	解析变量	模型一（Enter）方式				模型二（Forward Stepwise）方式			
		B	Wald	Sig.	Exp(B)	B	Wald	Sig.	Exp(B)
个人因素	年级	0.191	0.285	0.594	1.21				
	性别	-0.07	0.013	0.911	0.932				
	专业范畴	-0.099	0.03	0.863	0.905				
	高考地区	0.024	0.011	0.918	1.024				
	家庭户口类型	1.353	4.079	0.043**	3.868	1.234	6.389	0.002***	9.941
	经济条件	-0.4	1.459	0.227	0.67				
	加权平均绩点	0.693	4.693	0.03**	2.001				
	参与社会实践活动频率	-0.429	0.418	0.518	0.651				
	参与学术活动频率	-0.423	0.661	0.416	0.655				
	学习兴趣	0.418	0.91	0.34	1.519				
	对所学专业满意度	-0.559	1.685	0.194	0.572				
	对所在高校的满意度	0.063	0.028	0.867	1.065				
	是否了解准备读研相关信息	0.661	3.589	0.058*	1.937	0.584	4.889	0.027**	1.793
	对硕士学位证的竞争力影响的看法	0.743	2.823	0.093	2.103	1.118	11.019	0.001***	3.06
	对就业形势的态度	-0.814	3.94	0.047**	0.443				
政策因素	相关政策影响	0.616	3.314	0.069*	1.852				
社会因素	家人的期望	0.249	0.64	0.424	1.283				
	老师的期望	-0.051	0.019	0.89	0.95				
	同学或朋友的影响	0.639	3.08	0.079*	1.895	0.521	5.368	0.021**	1.684
	社会趋势	2.811	9.845	0.002***	16.623	2.292	9.696	0.002***	9.889
常量	常量	-12.043	11.627	0.001	0.000	-12.042	26.79	0.000	0.000
模型整体检验	预测准确率								
	-2对数似然值	94.748				113.184			
	Cox & Snell R 方	0.445				0.352			
	Nagelkerke R 方	0.594				0.47			

*、**、***分别表示在10%、5%和1%水平下显著

解 答 栏

这两个方法在某种程度上没有可比性，主要是看你想做什么。Enter 法是所有变量全部强制纳入的方法。它主要应用于解释型回归，即模型指标的选取是根据理论基础/参考文献/客观常识做的，是用来解释这些变量对因变量的影响作用的；而步进法（前向/后向）更多的是适用于预测型回归，即不是太关注各个自变量对因变量具体是怎样的影响，更多的是想探究哪些变量是显著影响因变量的，这个方法更多的是用在预测上的。所以就我个人而言，我的建议是楼主考虑清楚你想要的是什么，一般而言，你是基于基础/参考文献/客观常识做了变量的研究假设的，所以我推荐使用 Enter 法。如果变量间存在多重共线性，推荐主成分回归，也不推荐步进法。祝好运~



提 问：

求助：关于计量方程回归结果比较大小问题

求大神来指点下，

方程 1：Y=X1+A+B+C

方程 2：Y=X2+A+B+C

是不是 X1 和 X2 必须要取对数之后，前面的系数才能进行比较啊？？？

然后还有关于交互项的比较问题

方程 1：Y=X1+x1*g+A+B+C

方程 2：Y=X1+X1*m+A+B+C

这个方程比较 X1 通过 G 还是 M 的影响机制那个好，G 和 M 还需要取对数吗？

解 答 栏

首先，楼主可能需要明确的是，我们在构建模型时，为什么要取对数以及何时取对数？当变量呈现有偏态分布，存在极端异常值时，这时一般会通过取对数的方式去减缓数据的波动趋势，消除一些极端异常值对模型造成的影响；其次，关于楼主的第一个问题，如果 X1 与 X2 专业意义不一样，量纲不一样，那么就我个人看来，直接构建两个模型比较系数大小是无意义的。建议楼主将 X1 与 X2 纳入一个模型：Y=a1*x1+a2*2+u，然后比较 a1 与 a2 的标准化回归系数大小（标准化回归系数消除了变量的量纲影响），这样就能比较哪个变量对因变量的影响更大了；最后，楼主要通过方程比较 X1 通过 G 还是 M 的影响机制哪个好，直接看两个模型哪个拟合的更好，哪个交互项更显著，能很好解释中介过程就好的。至于取不取对数可能还是要回到数据的分布本身，而与其它无关。祝好运~



提 问：

谁知道哪本书里面有具体讲泊松回归负二项回归以及弹性分析的内容

谁知道哪本书里面有具体讲泊松回归负二项回归以及弹性分析的内容

解 答 栏

推荐陈强老师的《高级计量经济学及 STATA 应用》（第二版）一书，里面有相关章节。好像是因为版权的原因，论坛只有第一版的电子版。链接如下：http://bbs.pinggu.org/thread-3975548-1-1.html。祝好运~



提 问：

怎么做 robust 的 ols 呀？

怎么输出调整过稳健性异方差的回归结果。

与之对应的 STATA 命令是 reg y x，robust

解 答 栏

STATA 命令就是 reg y x，robust，使用这个命令后回归表格给出的标准误就是稳健标准误了（即消除了异方差后的标准误）。结果汇报时其它的不变，只是以前的标准误变成现在的稳健标准误了。



提 问：

关于概率的问题

我现在随机产生若干个随机变量，并且知道它的理论分布，我该如何通过图像来验证那？

不妨产生的是服从高斯分布的随机变量，如何判断这些随机变量确实服从高斯分布，请不要说验证方差，均值，

我想从更加明显的方法，通过图像给出直观的表达，谢谢！

求大神给个思路，再次感谢！

当然我们可以画出随机变量的直方图，但是如何与理论的分布进行对比，看是否符合那，求大神指点一二。

解 答 栏

如果楼主想验证的是高斯分布（正态分布），直接通过 PP 图和 QQ 图看即可，其它的分布应该也对应着有相应的图形。不过我个人认为图形只能辅助大致

的判断，要具体判断一个分布是何种类型，还是要结合统计量来看的。如判断一个变量是否符合正态分布，其可以通过 Shapiro-wilk 的 W 检验（sata 对应代码 swilk）和 Shapiro-Francia 的 W'检验（sata 对应代码 sfrancia）实现。祝好运~

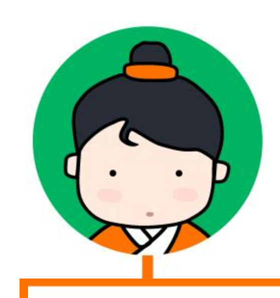
5.2

分类数据回归

1.1 logit/probit 模型说明

限值因变量（LDV）模型在计量经济分析中应用十分广泛，对于因变量取值范围明显受到限制的，便可考虑使用该方法。一般来说 logit 和 probit 模型是常用的两种限值因变量模型（被解释变量介于 0 到 1 之间），此外还有线性概率模型（LPM），但由于该模型拟合出来的概率可能小于 0 或大于 1，并且任何一个解释变量的偏效应都是不变的，因此在一定程度上限制了该模型的使用。

注：logistic 模型即 logit 模型，二者表示的含义一样，logit 模型中使用的函数是标准 logistic 随机变量的累计分布函数，因此有时也被称为 logistic 回归。



提 问：

logit 回归如何计算边际效应？用 SPSS 软件可以计算出来吗

logit 回归如何计算边际效应？用 SPSS 可以吗

解 答 栏

就我知道的而言，SPSS 是可以计算 logit 回归时的边际效应的。不过建议楼主先看看你模型的因变量是何种类型，是二分类、无序多分类还是有序多分类，类型确定了才好选择相应的 Logistic 回归模型，然后结合模型结果里面的 B（偏回归系数值）和 EXP(B)值对模型进行解释。可找相关文献看看如何解释。模型的原理和操作以及结果解读可参考张文彤老师《SPSS 统计分析教程》，链接如下：http://bbs.pinggu.org/thread-3573256-1-1.html。祝好运~

追 问：

非常感谢，因变量是二分类变量，我想计算 dy/dx 的值，不知怎么用 SPSS 处理，您能给我解答一下？

回　答：

SPSS 做 dy/dx 是默认取变量均值处的边际概率，如果想得到任意点的边际概率值，SPSS 我不知道如何实现。不过 STATA 的 margins 命令可以实现（可在 STATA 中使用 help margins 看看具体实现过程）。祝好运~



提 问：

logistic 回归如何将频数做为因变量

求教各位大神，我的问题如下：1、因变量为如下数据，在某个格网中，有两类家庭，可以计算出每类家庭出现的频数（示例如下）

网格编号	家庭类型	出现频数
1	1	90%
1	2	10%

2	1	30%
2	2	70%

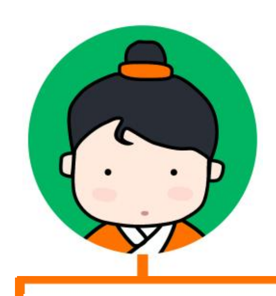
2、自变量为影响家庭出现在这个网格的因素，如土地价格、交通便利程度等值（0-100）

3、需要建立每类家庭出现的频数与影响因素之间的回归方程

该选择哪种回归模型？因变量需要做进一步处理吗？

解 答 栏

看了楼主的描述，貌似楼主还没有明确自己想做什么的。模型的选择是看因变量类型的。按照我个人的理解，你的因变量可以分为两种。一是家庭是否出现在网格中（这是二分类，可以用二元 Logistic 回归/probit 回归）；二是每类家庭出现的频数与影响因素之间的回归方程，这里的因变量是每类家庭出现的频数，是一个连续性的变量，可以做多元线性回归。如果楼主想做 logistic 回归，可人为的根据每类家庭出现的频数分组（如出现 70 次以上为高，40-69 为中，39 及以下为低，分类标准看你自己），然后建立有序多分类 logistic 回归构建模型即可。祝好运~



提 问：

多分类无序标量

向大家求助，多元线性回归，多分类无序变量怎么处理啊，要怎么处理，才能参与运算

解 答 栏

有点没明白楼主想表达的意思。其实，要构建什么样的模型，主要是取决于因变量类型的。当因变量为连续性的变量时，可构建楼主说的多元线性回归；当因变量为类别变量时（二分类、无序多分类，有许多分类），可构建（二分类、无序多分类，有许多分类）logistic 回归模型。建议楼主系统的学习下相关方法的原理再来构建模型吧。推荐张文彤老师《SPSS 统计分析教程》，链接如下：http://bbs.pinggu.org/thread-3573256-1-1.html。祝好运~



提 问：

logistic 回归

logistic 回归的方程怎么检验？如何看得到选出来的变量能解释多少的变异？

解 答 栏

楼主的问题实质是模型拟合优度的问题，logit 模型不像多元线性回归，会有一个 R^2，由于不存在平方和分解公式，故无法计算 R^2，但仍然会报告一个准 R^2(Pseudo R^2)，但这个作为一个参考就好；此外，就我个人所知道的，判断 logistic 回归模型拟合优度的另外一个方法是看其正确预测的百分比。祝好运~



提 问：

SPSS 用什么方法分析下列数据比较合适？（请进）

因变量：二分变量(是或否）

自变量：多选题(职业类型，忘记职业类型属于什么变量了- -，在 SPSS 中设置选中=1，未选中=0）想探究自变量对因变量有何影响？

解 答 栏

如果非要做模型分析，由于楼主的因变量为二分变量，可以考虑用二分类 Logistic 回归或者二分类 Probit 回归。PS:楼主的自变量职业类型是类别变量。祝好运~



提 问：

logistic 模型的问题

请问问卷中有不知道的选项做多元 logistic 模型怎么处理？

解 答 栏

可以考虑将其作为一个参照组，照常规的 logit 模型进行解释就行了。如果选择该项的个案不是太多，且样本量足够大，可以考虑删除这些个案。祝好运~



提 问：

SPSS 跑有序多分类 logistic 回归为什么没有 OR 值？

只有 estimate，怎么将 OR 值进行转换呢？

解 答 栏

SPSS 在构建有序多分类 logistic 回归模型时，不会报道 OR 值，不过会报道偏回归系数 B 值 ,可在 excel 里面计算 exp(B)得到 OR 值 ,因为 OR= exp(B)。祝好运~



提 问：请教几个 logistic 回归的问题：P for trend，标准回归系数

有几个关于 logistic 回归的问题请教各位大神：

1.自变量是连续变量时，将其用四分位数法分为四个水平，以第一水平为参照设哑变量，可以得到其余三水平相对于第一水平的 OR 值，请问该自变量的 P for trend 该如何在 SPSS 中计算？我曾见有人将该自变量的四水平作为连续变量进入方程，所得 P 值即为 P for trend，这样是正确的吗？

2.P for trend 在文献中每每提及，它的具体意义究竟是什么？研究中往往希望最高水平相对于最低水平，P 值小于 0.05，如果上述情况满足，但是此时 P for trend 大于 0.05，该如何解释该结果？反过来，如果 P for trend 小于 0.05，但各水平对于最低水平的 P 值却大于 0.05，又该如何解释该结果？

3.在 SPSS 中，如以上述方式处理连续自变量，在最后结果中可以看到其余三水平相对于第一水平的 OR 值和 P 值，在这些结果上面一行，还有一个 P 值，请问该 P 值的含义究竟是什么，如图中红圈所示：

		B	S.E,	Wals	df	Sig.	Exp (B)
步骤 1[a]	年龄	.128	.010	149.305	1	.000	1.137
	性别	-.136	.170	.634	1	.426	.873
	HDL离散			5.872	3	.118	
	HDL离散(1)	-.220	.229	.918	1	.338	.803
	HDL离散(2)	-.240	.223	1.158	1	.282	.786
	HDL离散(3)	-.561	.233	5.804	1	.016	.571
	常量	-8.169	.687	141.268	1	.000	.000

4.自变量是连续变量时，文献中常见 per SD change 的说法，即自变量改变一个标准差，请问如何在 SPSS 中得到自变量改变一个标准差的 OR 值？如果将自变量作为连续变量直接进入方程只能得到改变一个单位，而不是一个标准差的 OR 值。

解 答 栏

第一个问题：类别变量在构建回归（包括 logistic 回归）模型时，是要以哑变量的方式进入模型的，因为类别变量的编码是无意义的，不能进行加减乘除的（如有 4 个类别，编码为 1,2,3,4 和编码为 10,20,30,40 一样）。所以直接将多类别变量纳入模型是不正确的（二分类类别变量可以直接纳入模型）。类别变量的虚拟化处理推荐张文彤老师《SPSS 统计分析教程》相关章节，链接如下：http://bbs.pinggu.org/thread-3573256-1-1.html；

第二个问题，涉及到类别变量关注组和参照组间的比较和解释问题。楼主这个问题三言两语说不清楚。建议找相关文献看看如何解释类别变量对因变量的影

响。同时，推荐谢宇老师《回归分析》一书（链接：http://bbs.pinggu.org/thread-3065280-1-1.html），看看 Logit 回归这章，对 Logit 模型结果的解释有好几种（如 OR，边际效应等）；

第三个问题，在我个人看来这个表格的模型构建方式是有问题的，做的人没有将类别变量处理为哑变量纳入模型，而是直接将多类别变量纳入模型，所以才会出现既有多类别变量的估计，又有哑变量的估计。我个人觉得正确的做法是利用 recode 命令将多类别变量处理为哑变量，然后直接将哑变量纳入模型建模，就只会得到关注组相对于参照组的结果(哑变量的处理方式具体可参见第一个问题的回答)。

最后一个问题，就我个人所知道的而言，大家在对 logistic 回归模型进行解释时，可能更多的还是关注偏回归系数 B 的大小和方向以及 OR 值（比数比），对标准差的变化关注可能较少。祝好运~

5.3

时间序列

5.3.1 时间序列

时间序列作为计量经济学研究的主要范畴之一，被广泛使用在各个领域，特别是 Granger 提出协整理论后，时间序列为克服虚假回归，提高模型估计质量提供了重要的理论支撑。由于时间序列研究问题较多，方法发展也十分迅速，这里仅对论坛中较为普遍的问题答疑进行汇总。

1.1 平稳时间序列模型

时间序列的平稳性是研究时间序列问题的基础。

推荐阅读：

[经验分享] 开学大献礼：怎样理解时间序列的“平稳性”？

单位根的检验方法较多，主要以 ADF 检验和 PP 检验为主。

推荐阅读：

[经验分享] 用 eviews 做单位根检验的方法

5.3.2 ARMA 模型

ARMA 模型作为单变量时间序列模型，尽管提出较早，但其在短期内预测性能的优势使得该方法至今仍是建模的主要方法之一。

推荐阅读：

【时间简“识”】6.千辛万苦终于走上了 ARMA 之路

【时间简“识”】4.开启 ARMA 之旅——AR 篇

【时间简“识”】5.开启 ARMA 之旅——MA 篇

【时间简“识”】7.ARMA 路障清除之确定性分析

5.3.3 VAR 模型

向量自回归模型自 1980 年 Sims 提出以来，一直被众多学者接受和使用，在国内使用该方法所做的论文不计奇数。不仅因为该方法结果较为理想，同时也因为该方法容易上手。

推荐阅读：

STATA 基本操作汇总——时间序列专项

VAR 模型的演变、最新发展及其应用

5.3.4 协整理论

协整理论的发展为现代时间序列计量经济的发展提供了坚实的基础。协整检验方法近些年来也在逐步演变和发展，出了传统时间序列中的 EG 两步法，JJ 检

验等，也有了新的拓展和应用。

阅读推荐：

[Eviews] 〖素质笔记〗Eviews 8 新功能之三——面板协整

5.3.5 ARCH 模型

ARCH 族模型经过多年的拓展，发展已经较为完善，作为研究金融数据波动聚集性方面获得了较为满意的结果。

推荐阅读：

GARCH 模型的应用总结



提 问：

如何提取特征值，建立什么模型？

如题。我对一组样本（100 个）做了为期三十天的监测（每天测一次，样本的颜色随着时间在发生变化，读到的颜色数据有 L、a、b 三种），现在想借助这个三个数据与时间 t 建立一个预测模型。

1、这三个数据并非都随着时间的增加与样本呈现出明显的相关，其中 a 值得相关性较好，L 的就很乱，但整体趋势上都有联系。所以我想知道应该用什么方法去筛选出特征值。（逐步回归？主成分分析？）

2、为什么对 a 值取自然对数后，lna 与时间 t 的关系就变成了近似线性的？

3、我应该建立线性的模型还是用决策树或者神经网络？

L、a、b 三个值随时间的变化如下图。因为之前从未学过统计学，自学了 SPSS 个把月但仍然没有头绪，谢谢各位大神帮忙解答一下。

解 答 栏

楼主这套数据是时间序列数据，需要对数据的不同期（即不同天）间是否相关做相应的检验，然后做了相应的处理后才做下一步分析。建议楼主找时间序列相关的资料看看，然后再进一步做分析。祝好运。

举个例子，国家今年颁布了一项政策，除了对今年影响外，明年后年甚至以后很多年可能都有影响。这就是政策效应在不同年间的相关。同理，用到你这里这个时间序列样本，不知道会不会有存在这种影响（因为我不懂你这个专业方面的东西，所以我只能先猜测有关）。序列相关是时间序列数据的特点之一，一般会先对其进行检验处理，再做进一步分析。



提 问：

面板 logit 模型回归中想得到经过异方差修正的标准误怎么做？

各位大神，本然在做面板 logit 后得出一个 Z 统计量，想知道这个 Z 统计量指的是什么？

我想得到经过异方差修正后的标准误，我加入了 vce(robust),但是 STATA 提示错误，不知道怎么回事！

解 答 栏

Z 统计量与 t 统计量和 F 统计量类似，是判断模型偏回归系数与 0 之间差异是否显著的统计量。如果你在构建模型前，对数据做了预处理，使得变量分布基本符合正态分布，不含极端异常值且自变量间不存在严重多重共线性，那结果基本就是稳健的。如果非要加稳健标准误，用 vce(cluster)命令，而非你的 vce（robust）命令。

追　问

很感谢你的回复，我试过了 vce(cluster)也是不行的，而且我还是有疑问，就是在图片中的 vce（robust）的解释是 huber-white 异方差修正后的标准误，而且我看了很多文献他们也是用的异方差修正后的标准误，我的为什么不能用，很疑惑，希望能够得到你的解答。

回　答

这里的 cluster 是模型变量的 id 对应的那个变量 ,而非直接在命令语句中写上 vce(cluster)。其实，就像我上条回复说的，对原始数据做了预处理后结果基本是稳健的，就不用太担心异方差的问题。祝好运~



提 问：

关于 panel data 的分析

请问关于面板 data 分析主要集中在那些部分 那个 fix effect 和 random effect 在解释回归方程的时候应该怎么解释呢

解 答 栏

线性面板回归常用的方法为混合 OLS 回归、固定效应模型和随机效应模型。具体选择何种方法估计模型需要根据相关的检验统计量来判定。所以我的建议是楼主先把这三种方法的原理、优劣势和软件实现搞清楚了，其结果解释自然就不存在问题了。推荐陈强老师《高级计量经济学及 STATA 应用》一书，有原理、代码和案例介绍。链接：http://bbs.pinggu.org/thread-3812173-1-1.html。祝好运~



提 问：

纵向数据分析方法有哪些？

响应变量有重复测量数据，对应自变量也有重复测量数据，这种纵向数据分析有什么方法吗？

解 答 栏

楼主的数据类型是面板数据，故而用面板回归即可。不过，方法的选择要视具体的因变量类型而定。如果因变量是连续性变量，那一般会在混合 OLS 回归、固定效应模型和随机效应模型间进行方法选择，有相应的检验统计量会协助判断；如果因变量是类别变量，那就用面板 logit 模型即可（对应着有二分类、有序多分类和无序多分类三种）。推荐陈强老师《高级计量经济学及 STATA 应用》一书，有以上介绍方法的原理介绍、代码实现和案例分析。链接：http://bbs.pinggu.org/thread-3989052-1-1.html。祝好运~

5.4

一级检验

5.4.1 一级检验：T、F 检验&可决系数 R^2

如何理解可决系数 R^2

可决系数也被称作 R^2，是测定多个变量间相关关系密切程度的统计分析指标，它也是反映多个自变量对因变量的联合的影响程度。可决系数越大,自变量对因变量的解释程度越高，自变量引起的变动占总变动的百分比高。观察点在回归直线附近越密集。可决系数的取值范围在 0 到 1 之间，它是一个非负统计量。随着抽样的不同而不同，既是随样本而变动的统计量。

特点：

（1）可决系数是非负的统计量

（2）可决系数的取值范围：$0<=R^2<=1$

（3）可决系数是样本观测值的函数，可决系数 R^2是随机抽样而变动的随机变量。为此，对可决系数的统计可靠性也应进行检验。

1、T、F 检验简介

T 检验，亦称 student t 检验（Student's t test），主要用于样本含量较小（例如 n<30），总体标准差σ未知的正态分布资料。t 检验是用 t 分布理论来推论差异发生的概率，从而比较两个平均数的差异是否显著。它与 Z 检验、卡方检验并列。

在回归分析中，我们的系数检验就会采用 t 检验，在应用时，以下几点大家一定要加以重视：

（1）选用的检验方法必须符合其适用条件(注意：t 检验的前提是资料服从正态分布）。理论上，即使样本量很小时，也可以进行 t 检验。（如样本量为 10，一些学者声称甚至更小的样本也行），只要每组中变量呈正态分布，两组方差不会明显不同。如上所述，可以通过观察数据的分布或进行正态性检验估计数据的正态假设。方差齐性的假设可进行 F 检验，或进行更有效的 Levene's 检验。如果不满足这些条件，只好使用非参数检验代替 t 检验进行两组间均值的比较。

（2）区分单侧检验和双侧检验。单侧检验的界值小于双侧检验的界值，因此更容易拒绝，犯第 I 错误的可能性大。t 检验中的 p 值是接受两均值存在差异这个假设可能犯错的概率。在统计学上，当两组观察对象总体中的确不存在差别时，这个概率与我们拒绝了该假设有关。一些学者认为如果差异具有特定的方向性，我们只要考虑单侧概率分布，将所得到 t-检验的 P 值分为两半。另一些学者则认为无论何种情况下都要报告标准的双侧 t 检验概率。

（3）假设检验的结论不能绝对化。当一个统计量的值落在临界域内，这个统计量是统计上显著的，这时拒绝虚拟假设。当一个统计量的值落在接受域中，这个检验是统计上不显著的，这是不拒绝虚拟假设 H0。因为，其不显著结果的原因有可能是样本数量不够拒绝 H0 ，有可能犯第 I 类错误。

（4）正确理解 P 值与差别有无统计学意义。P 越小，不是说明实际差别越大，而是说越有理由拒绝 H0 ，越有理由说明两者有差异，差别有无统计学意义和有无专业上的实际意义并不完全相同。

（5）假设检验和可信区间的关系结论具有一致性差异：提供的信息不同区间估计给出总体均值可能取值范围，但不给出确切的概率值，假设检验可以给出 H0 成立与否的概率。

F 检验又叫方差齐性检验。在两样本 t 检验中要用到 F 检验。从两研究总体中随机抽取样本，要对这两个样本进行比较的时候，首先要判断两总体方差是否相同，即方差齐性。若两总体方差相等，则直接用 t 检验，若不等，可采用 t'检验或变量变换或秩和检验等方法。其中要判断两总体方差是否相等，就可以用 F 检验。若要更详细的了解 F 检验，我们可以看后续的 4.3 方差分析篇章。



提 问：

SPSS 里进行回归分析时控制变量的 sig 大于 0.05，这样可以继续进行分析吗？

系数[a]

模型		非标准化系数 B	非标准化系数 标准 误差	标准系数 试用版	t	Sig.
1	(常量)	.148	.038		3.925	.000
	长期负债率	.641	.154	.307	4.168	.000
	短期负债率	-.496	.165	-.404	-3.017	.003
	银行借款率	.325	.176	.206	1.847	.066
	商业信用率	.519	.188	.299	2.763	.006
	现金流量净额标准化	-.022	.147	-.008	-.151	.880
	营业收入标准化	.004	.021	.011	.211	.833

a. 因变量：投资-资本存量

最下面的现金流量和营业收入是控制变量，加入以后自变量都是显著的，可是控制变量数特别大，不小于 0.05 的显著性，这样可以继续进行分析写论文吗？我可以 就说在加入控制变量以后自变量和因变量是显著的这样行吗？我第一次用 SPSS 很多东西不太明白~请教一下各位~

解 答 栏

首先，建议楼主想想在构建模型时为什么要选择这些变量，是基于理论、文献还是根据常识想的？其次，要看楼主要做什么。回归分为解释型回归和预测型回归。目的不一样，结果保留就会不一样（解释型回归和预测型回归差别和案例可看邱皓政老师《量化研究与统计分析——SPSS 数据分析范例解析》第十一章）；最后，针对楼主这个问题来说，如果是基于理论或者文献选了这些指标，且加入最后两个不显著的指标后其它指标才显著，那么在最终构建模型时可以加入最后两个指标，并在专业上寻求其不显著的原因。同时，在解释其它显著的指标时，说明这是在控制其它变量情况下（即控制了这两个不显著指标情况下）得到的结果：自变量每变化一单位，因变量怎么变化；如果不加最后两个指标，其它变量也显著，且这两个指标也并非在专业上来说非加入不可的，那么可以不用加入。

追 问

我是想研究企业负债对投资行为的影响~然后是看了很多文献跟论文基本上都选了这个 2 个控制变量~可是我刚才照着您分方法剔除控制变量以后自变量和因变量之间也是显著的结果是一样的。这样的话那我是不是可以不需要控制变量了呢？

回　答

如果情况是你描述的这样。那么可分情况讨论，如果这两个控制变量是严格外生的，那么在你这个模型里面可以删掉，因为删了它们也不会对模型结果造成什么影响（如遗漏变量偏误）；如果这两个变量删了后会造成模型遗漏变量偏误，那么即使它们不显著，那还是把它放在模型里面吧。祝好运~



提 问：

回归系数的变化趋势与 SIG 值的变化趋势

请教各位实证大牛们，论文中的多元回归模型 1 如：Y=AX1+BX2+CX3+DX4 （X1 为解释变量，X2-X4 为控制变量），将数据导入 SPSS 中的运行结果显示，X1 的标准系数=0.097,sig 值=0.025　　模型 2:Y=AX1+BX2+CX3+DX4+FX1*X2(模型 1 与模型 2 中的所有变量都是一样的，模型 2 在模型的基础上加入了 X1*X2 的交互项)，模型 2 的 SPSS 回归结果

显示 X1 的标准系数=0.091,sig 值=0.077 请问：解释变量 X1 的标准回归系数由模型 1 的 0.097 下降到了模型 2 的 0.091，其对应的 SIG 值由模型 1 的 0.025 上升到了模型 2 的 0.077,同样的一个变量的回归系数下降的幅度与其对应的 SIG 值上升的趋势是否存在严格的倍数关系，例如本文中解释变量的系数由模型 1 到模型 2 中下降了 0.006，而其对应的 SIG 值从模型 1 到模型 2 上升了 0.052（变化比=0.052/0.006=8.67），这样的结果符合统计学规律吗？还是操作有误导致？或者是说变量系数的变化多少与其对应的 SIG 值的变化幅度之间所程的比例没有一个严格的范围，出现什么样的结果都是正常的？

解 答 栏

推荐楼主仔细看看谢宇老师《回归分析》交互项这章（链接：http://bbs.pinggu.org/thread-3065280-1-1.html），里面详细了阐明了交互项的使用前提和适用条件，以及如何解释。一般而言，X1，X2 与 Y 单独构建模型时均显著时，才会考虑构建 x1*x2。同时，为了避免严重多重共线性的影响（x1,x2 与 x1*x2 间），需要对变量进行中心化处理然后再构建交互项，并进一步做回归分析。

追　问：

谢谢您的回复，关于加入交互项的前提和适用范围我比较清楚，我的问题主要是本文中解释变量的系数由模型 1 到模型 2 中下降了 0.006，而其对应的 SIG 值从模型 1 到模型 2 上升了 0.052（变化比=0.052/0.006=8.67），这样的结果符合统计学规律吗？您看能帮我解释一下吗？谢谢!

回　答：

我看楼主这个操作看不出来有没有对 x1 与 x2 先做中心化处理，然后再构建中心化后的 x1*x2 去构建模型。依据 y=a1+b1*x1+b2*x2+b3*x1*x2+u，对 x2 求偏导得到其偏导数=b2+b3*x1，所以加入交互项后其系数的变化是与 x1 的变化相关的。我的理解是其对应的 sig 值不会有固定的比例变化。我也觉得你看这个 sig 值的比例变化没多大实际意义。加入交互项后更应该看的是做了交互项后对模型截距和斜率的影响，探究这个更有意义。

追　问：

谢谢你，我会按照你的提示去，把那本回归分析中的关于交互项的内容好好学习一下。我们是学财务管理的硕士，虽然开设了计量经济学和多元统计分析等课程，但是真的比不上统计科班出身的学生 我们都只会依葫芦画瓢，很难理解其中的原理和一些该有的处理，而且我们专业的绝大部分导师对于实证分析也是半桶水，所以我们都以为某个变量的回归系数的变化趋势与其对应的 SIG 值的

变化有一个大概的比例，不能超过某个比例？

回　答：

这个问题比较复杂，需要去搞清楚回归系数对应的 t 值和 sig 值的背后计算过程。需要深厚的线性代数和概率论功底。所以也不必太深究，懂原理会软件操作和结果解读就行。你可以参考下相关文献的处理方式的。祝好运~



提 问：

t 分布当中自由度，是源于什么定义的

请高手解释一下 t 分布自由度的含义

解 答 栏

根据谢宇老师《回归分析》(链接：

http://bbs.pinggu.org/thread-3065280-1-1.html）一书中的介绍：自由度是通过样本统计量来估计总体参数时必须涉及的一个基本概念，指的是计算样本统计量时能自由取值的数值个数。当做 t 检验时，是用样本方差去对总体方差进行估计。需要变量减去观测样本的均值，故而样本中只有 n-1 个自由取值。确定了 n-1 个数，基于均值，第 n 个数就确定了。所以一般来讲，丧失的自由度数目也就是需要估计的参数的数目，或者约束条件的数目。祝好运~



提 问：

只有均值，如何分析?

试验时遇到实验组与对照组数据只有均值，最大值和最小值这三类，请大神们指教，可以用什么统计方法分析？

解 答 栏

由于楼主给的信息不详细，所以现在只能宽泛的给点建议：如 t 检验、方差分析等。具体用哪种 t 检验，什么方差分析方法需要进一步详细的信息，如你的研究设计大概怎样，你想得到什么样的结果等。

追　问

谢谢指点，因为搜集到的都是均值数据，总共 3 组数据之间进行比较差异，每组 10 个均值数据，把所有均值当成连续性数据进行方差分析。不知是否恰当，是否有更好的方法？

回　答

照你这么说，就用方差分析即可。不过方差分析又要分很多种（如单因素方差分析，多因素方差分析），建议找本书来看看具体要用哪种方差分析的。推荐张文彤老师《SPSS 统计分析教程》。链接：

http://bbs.pinggu.org/thread-3573256-1-1.html。祝好运~



提 问：

截面数据的系数有什么意义吗？

前段时间接了一个烂尾项目，13 年 X 市的公园保安配置的项目。前期的专家调研得出 4 个影响因素，将保安划分为固定岗和巡逻岗，然后就选取了 30 多家公园进行问卷调查。收集完的数据就是当年各个公园“的影响因素以及当时的保安人数与实际需求人数”的数据，问题就是将这 30 几个公园数据进行回归，得出的结论是完全不显著的，而且现在已经无法再去收集公园的数据了，只有现有的数据作为基础。能有什么办法将这个结果挽救的吗？我的处理是先检验异方差，然后用加权最小二乘法消除，可是我这个是多元的回归，要怎么处理呢？

解 答 栏

相对于你这套横截面数据来说，方法选对了，然后极端异常值处理了，得到的结果可能就基本没问题了。偏回归系数不显著可能需要从专业层面或者现实角

度去进行解释了。你这个问题如果要用 SPSS 软件去处理，稍微麻烦点，可能还要去探究具体的异方差形式，然后才能确定权重，进而构建加权最小二乘法。我的建议有二：

一是你看看你这个领域相关文献，别人的异方差权重是怎么求的；

二是换 STATA/eviews/sas 等软件，后面加个 robust 稳健估计，用稳健标准误去构建统计量。这个方法的好处是不用管异方差形式，直接就得到保守的估计结果。PS：如果横截面数据预处理得当（如正偏态分布数据取对数后会在一定程度上消除异方差），一般不会存在异常差的。祝好运~



提 问：

单侧检验与双侧（双尾）检验有什么区别？适用范围分别是什么呢？

单侧检验与双侧（双尾）检验有什么区别？适用范围分别是什么呢？

解 答 栏

假设检验可以进一步分为单尾检验（也称单侧检验，one-tailed test）和双尾检验（也称双侧检验，two-tailed test）。

单尾检验是指否定域在曲线的左端或右端区域的情况，双尾检验是指否定域在曲线的两端区域的情况，如图 2－11 和图 2－12 所示。

一个检验是双尾还是单尾取决于对应于零假设的备择假设 H_1（研究假设）。在单尾检验中，可选任一方向的单侧备择假设：如果选 H_1：$\theta<\theta_0$，则称此单尾检验为左侧检验；如果选 H_1：$\theta>\theta_0$，则称此单尾检验为右侧检验。在双尾检验中，备择假设是无方向或双向的 H_1：$\theta\neq\theta_0$。

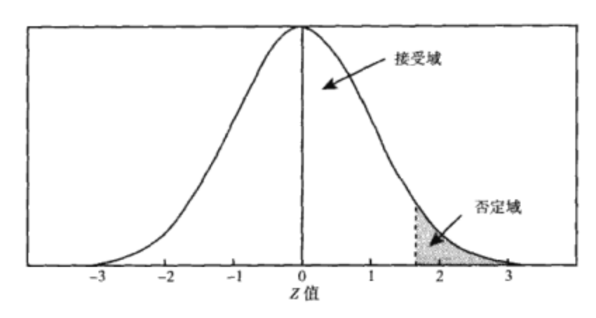


图 2－11　单尾检验（右侧）

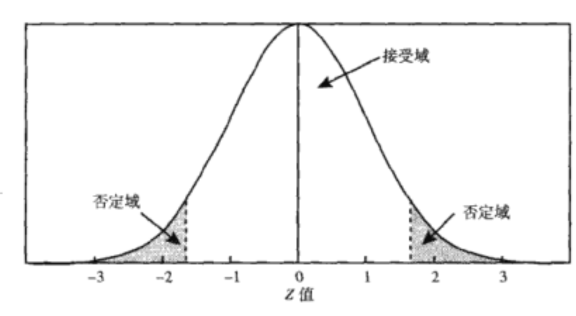
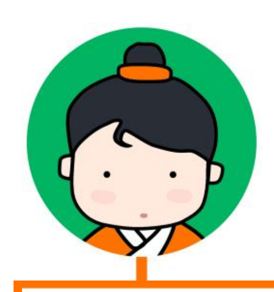


图 2－12　双尾检验

以上是谢宇老师《回归分析》一书中的内容，希望对你有帮助。

http://bbs.pinggu.org/thread-3065280-1-1.html，这是这本书的下载地址。



提 问：

回归系数和显著性检验

用 STATA 做动态面板回归，得出解释变量系数 a1、 a2、 a3，现在要进行 a1+a2+a3 是否显著大于零的检验，如何操作，谢谢赐教！

解 答 栏

先用 egen/gen 命令产生一个新的变量，使其等于 a1+a2+a3。如 gen a=a1+a2+a3，然后用 test a 命令即可。祝好运~

5.4.2 一级检验：P 值

P 值（P value）就是当原假设为真时所得到的样本观察结果或更极端结果出现的概率。如果 P 值很小，说明原假设情况的发生的概率很小，而如果出现了，根据小概率原理，我们就有理由拒绝原假设，P 值越小，我们拒绝原假设的理由越充分。总之，P 值越小，表明结果越显著。但是检验的结果究竟是“显著的”、“中度显著的”还是“高度显著的”需要我们自己根据 P 值的大小和实际问题来解决。

一般情况下：

- 如果 $P<0.01$，说明是较强的判定结果，拒绝假定的参数取值。
- 如果 $0.01<P$ 值 <0.05，说明较弱的判定结果，拒绝假定的参数取值。

如果 P 值 >0.05，说明结果更倾向于接受假定的参数取值。



提 问：

用标准误怎么解释回归系数

回归之后，用标准误怎么解释系数结果？比如，0.587（0.057），前者为估计系数，后者为标准误。标准差和标准误一样么？我也看到过标准化系数可以解释为 X 变动一个单位的标准差对 Y 的影响，这里所谓的一个单位标准差就是指显示出来的这个标准误吗？不懂呢，希望知道的人告诉一下啊

解 答 栏

楼主所说的其实是标准化回归系数，而非标准误。标准差和标准误是有区别的。前者是方差开根号。或者是在标准差基础上除以根号 n 得到。不要混淆了。祝好运~

5.5

二级检验

5.5.1 二级检验：自相关&异方差

5.5.1.1 异方差

异方差的定义：

异方差性是相对于同方差而言的。所谓同方差，是为了保证回归参数估计量具有良好的统计性质，经典线性回归模型的一个重要假定：总体回归函数中的随机误差项满足同方差性，即它们都有相同的方差。如果这一假定不满足，即：随机误差项具有不同的方差，则称线性回归模型存在异方差性。

1、异方差的表现形式：

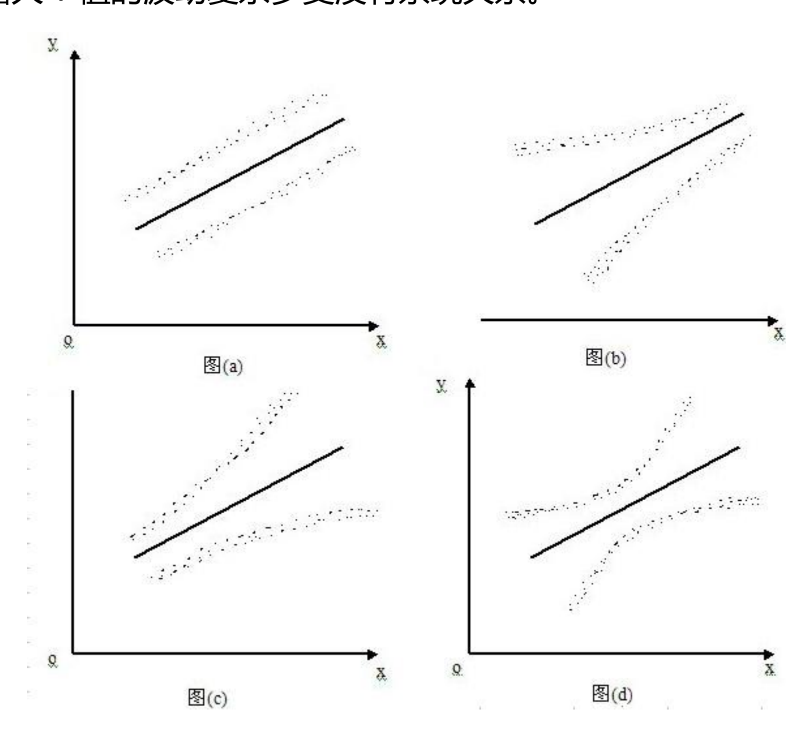
异方差性的几何直观表示形式，可借助观测值的散布图表示。以一元线性回归为例，在散布图上，就是样本残差平方随解释变量的变化而变化。

异方差一般可归结为三种类型：

（1）单调递增型：随 X 的增大而增大，即在 X 与 Y 的散点图中，表现为随着 X 值的增大 Y 值的波动越来越大

（2）单调递减型：随 X 的增大而减小，即在 X 与 Y 的散点图中，表现为随着 X 值的增大

（3）复杂型：与 X 的变化呈复杂形式，即在 X 与 Y 的散点图中，表现为随着 X 值的增大 Y 值的波动复杂多变没有系统关系。


图(a) 图(b) 图(c) 图(d)

2、异方差产生的原因：

1)模型中遗漏了某些解释变量:如果模型中只包含所要研究的几个主要因素，其他被省略的因素对被解释变量的影响都归入了随机误差项，则可能使随机误差项产生异方差性。

2）模型函数形式的设定误差:在一般情况下，解释变量与被解释变量之间的关系是比较复杂的非线性关系。在构造模型时，为了简化模型，用线性模型代替

了非线性关系，或者用简单的非线性模型代替了复杂的非线性关系，造成了模型关系不准确的误差。如将指数曲线模型误设成了线性模型，则误差有增大的趋势。

3）样本数据的测量误差:一方面，样本数据的测量误差常随时间的推移而逐步积累，从而会引起随机误差项的方差增加。另一方面，随着时间的推移，抽样技术和其他收集资料方法的改进，也使得样本的测量误差逐步减少，从而引起随机误差的方差减小。因此，在时间序列资料中，由于在不同时期测量误差的大小不同，从而随机项就不具有同方差性。

4）随机因素的影响:经济变量本身受很多随机因素影响（比如政策变动、自然灾害或金融危机等），不具有确定性和重复性，同时，社会经济问题涉及人的思维和行为，也涉及各阶层的物质利益，人的行为具有很多不确定因素。

因此，经济分析中经常会遇到异方差性的问题。而且经验表明，利用横截面数据建立模型时，由于在不同样本点上（解释变量之外）其他因素影响的差异较大，所以比时间序列资料更容易产生异方差性。在实际经济计量分析中，绝对严格的同方差性几乎是不可能的，异方差性可以说是一种普遍的现象。

如果出现异方差情况，它会导致：参数估计量仍然是线性无偏的，但不是有效的；异方差模型中的方差不再具有最小方差性；t 检验失去作用；模型的预测作用遭到破坏。

3、异方差的检验方法：

1）图示检验法。

①相关图分析。方差为随机变量的离散程度，通过观察 y 和 x 的相关图，可以观察的离散程度和解释变量之间的相关关系。若随 x 的增加，y 的离散程度呈逐渐增加或减少的趋势则表明模型存在着递增或者递减的异方差性。

②残差图分析。通过对模型残差分布的观察，如果分布的离散程度有明显扩大的趋势，则表明存在异方差性。图示检验法只能较简单粗略判断模型是否存在着异方差性。

2）Goldfeld - Quandt 检验法。将解释变量排序，分成两个部分利用样本 1 和样本 2 分别建立回归模型，并求出各自残差平方和。为在检验过程中“夸大”差异性，在样本中去掉 c 个样本数据（c= n/4）则构造 F 统计量，对于给定显著水平，根据 P 值判断。

3）怀特(white) 检验。White 检验是通过建立辅助回归模型的方法来判断异方差性。假设回归模型为二元线性回归模型，则 White 检验的步骤为：估计回归模型，计算残差；估计辅助回归模型：即将残差平方关于解释变量的一次项，二次项和交叉乘积项进行回归；计算辅助回归模型的判断系数，根据 P 值判定。

4）帕克检验和格里奇检验。通过建立残差序列对解释变量的辅助回归模型，判断随机项的误差和解释变量之间是否有较强的相关关系，以此来判断模型是否存在异方差性。

4、异方差的修正：

1）对模型变换，当可以确定异方差 的具体形式时，将模型作适当变换有可能消除或减轻异方差的影响。

2）加权最小二乘法，对原模型变换的方法与加权二乘法实际上是等价的，可以消除异方差。

3）模型的对数变换，运用对数变换能使测定变量值的尺度缩小。它可以将两个数值之间原来 10 倍的差异缩小到只有 2 倍的差异。其次，经过对数变换后的线性模型，其残差 e 表示相对误差，而相对误差往往比绝对误差有较小的差异。

5.5.1.2 自相关

自相关的定义：

如果随机误差项的各期望值之间存在着相关关系，这时，称随机误差项之间存在自相关性或序列相关。

2、自相关性产生的原因：

线性回归模型中随机误差项存在序列相关的原因很多，但主要是经济变量自身特点、数据特点、变量选择及模型函数形式选择引起的。

1）经济变量惯性的作用引起随机误差项自相关

2）经济行为的滞后性引起随机误差项自相关

3）一些随机因素的干扰或影响引起随机误差项自相关

4）模型设定误差引起随机误差项自相关

5）观测数据处理引起随机误差项序列相关

3、自相关的后果：

线性相关模型的随机误差项存在自相关的情况下，用 OLS（普通最小二乘法）进行参数估计，会造成以下几个方面的影响。

从高斯-马尔可夫定理的证明过程中可以看出，只有在同方差和非自相关性的条件下，OLS 估计才具有最小方差性。当模型存在自相关性时，OLS 估计仍然是无偏估计，但不再具有有效性。这与存在异方差性时的情况一样，说明存在其他的参数估计方法，其估计误差小于 OLS 估计的误差；也就是说，对于存在自相关性的模型，应该改用其他方法估计模型中的参数。

1）自相关不影响 OLS 估计量的线性和无偏性，但使之失去有效性

2）自相关的系数估计量将有相当大的方差

3）自相关系数的 T 检验不显著

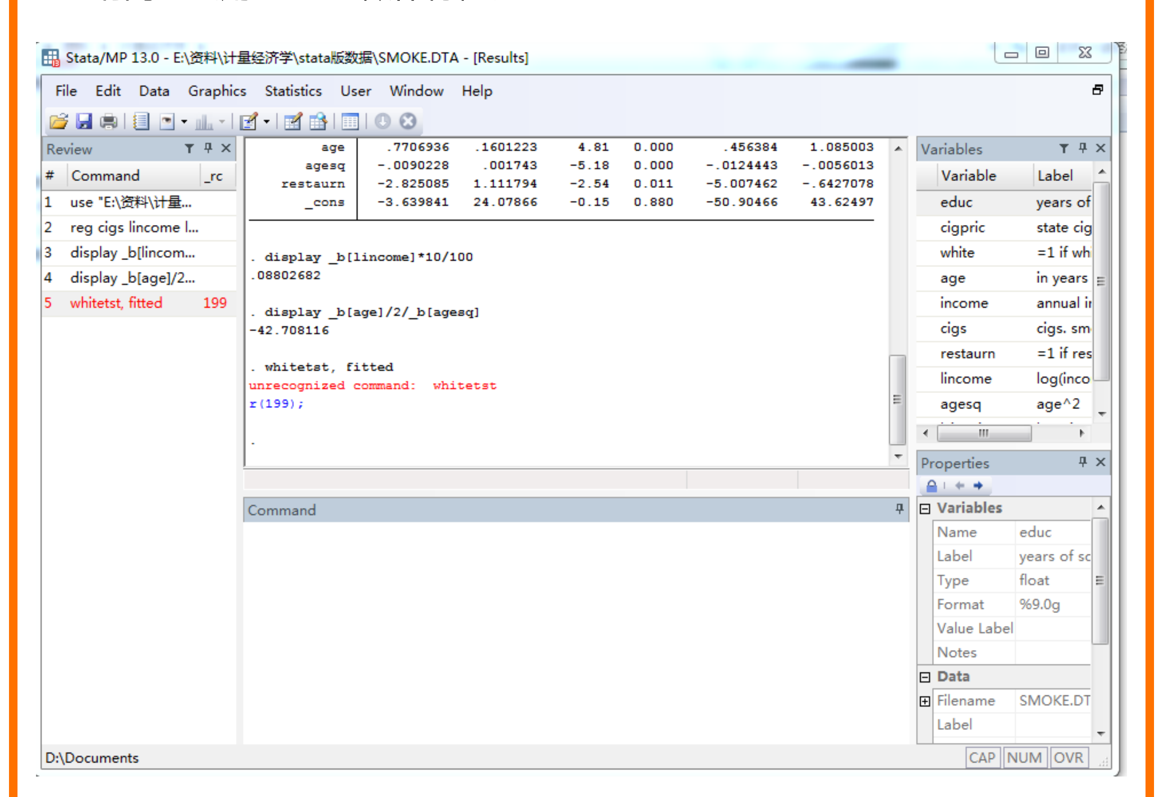
4）模型的预测功能失效



提 问：

请问怎么用 STATA 做怀特检验？

请问怎么用 STATA 做怀特检验？



到了图片中的那个步骤就错了！

解 答 栏

应该是软件还没有安装相关命令，用 findit whitetst 或者 ssc install whitetst 去安装命令后，再做检验。祝好运~



提 问：

STATA 下 WLS 的命令问题

请教一下各位，我做一个回归，文献里讲“weights are proportional to square root of N"，在这种做法下，我 STATA 下面进行 WLS，需要在 reg xxx，xxx ,xxx 后边加的是 aw=N ,还是 aw=根号 N？这个 STATA 里的 aw/aweight 命令是怎么操作的呀？谢谢。

解 答 栏

这个要确认权重函数（即楼主语句中 aw=的部分），而权重函数的确定首先要去检验模型中是哪个或哪几个变量存在异方差，然后还要去检验异方差的形式，确定好后再利用残差去求 aw=的部分。其实，如果模型存在异方差，一个一劳永逸的方式是直接用 robust 得到稳健标准误，而不要考虑异方差形式。祝好运~



提 问：

STATA 时间序列数据自相关的修正可以用 prais 语句吗

STATA 新手，在一篇文章上看到用 prais 语句，但是在其他地方并没有看到过 prais 语句，所以不太明白 prais 语句的具体适用范围，还请各位帮帮忙

解 答 栏

可以用 help prais 看看帮助文件呢。这个命令是解决时间序列一阶自相关的命令。祝好运~



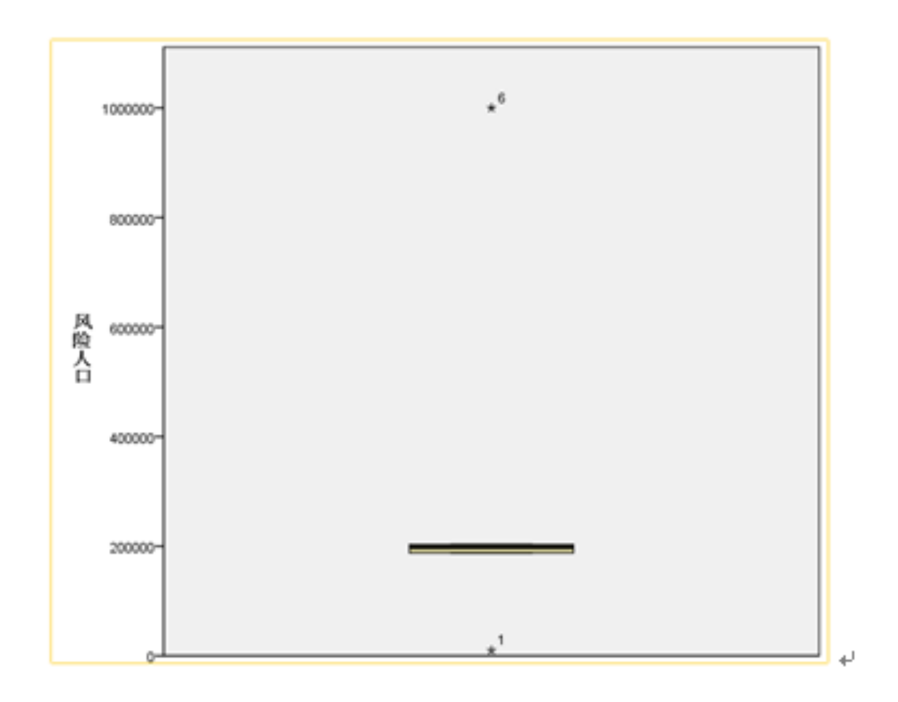
提 问：

线性回归检测异常值使用问题

本人最近在做温室气体数据统计分析工作，温室气体排放量统计有计算公式，即温室气体排放总量等于消耗量乘以相关系数再相加即可，请问能用线性回归的分析方法，用消耗量做因变量预测温室气体排放总量，再利用预测结果和真实结果采用残差和 COOK 距离的方法找异常值吗？

解 答 栏

异常值的寻找不用这么复杂的，直接使用 SPSS 做描述性统计分析：图形—图表构建器（弹出对话框，点确认）—选择箱图即可：



如上图，我们看到这个箱形图的上下方各有一个“*”，表明存在极端异常值。具体操作过程推荐楼主看看我个人写的实证分析中如何使用描述性统计分析寻找极端异常值，链接：http://bbs.pinggu.org/thread-3651449-1-1.html。祝好运~

5.5.2 多重共线性

1、多重共线性的定义：

多重共线性是指线性回归模型中的解释变量之间由于存在精确相关关系或高度相关关系而使模型估计失真或难以估计准确。一般来说，由于经济数据的限制使得模型设计不当，导致设计矩阵中解释变量间存在普遍的相关关系。完全共线性的情况并不多见，一般出现的是在一定程度上的共线性，即近似共线性。

2、产生的原因：

（1）经济变量相关的共同趋势

（2）滞后变量的引入

（3）样本资料的限制

3、多重共线性会造成什么影响：

（1）完全共线性下参数估计量不存在；

（2）近似共线性下 OLS 估计量非有效：多重共线性使参数估计值的方差增大，1/(1-r2)为方差膨胀因子（Variance Inflation Factor, VIF）;

（3）参数估计量经济含义不合理；

（4）变量的显著性检验失去意义，可能将重要的解释变量排除在模型之外；

（5）模型的预测功能失效。变大的方差容易使区间预测的“区间”变大，使预测失去意义。

需要注意：即使出现较高程度的多重共线性，OLS 估计量仍具有线性性等良好的统计性质。但是 OLS 法在统计推断上无法给出真正有用的信息。



提 问：

如何根据皮尔逊相关系数结果去掉有重复性测量嫌疑的指标

做一个关于指标遴选的课题，为了验证初选指标的之间的相互独立性，避免指标之间相关性过强而出现重复测量的现象，在进行因子分子之前先做 pearson 确定相关性。结果显示，部分指标间有显著性或非常显著性差异，有重复测量的嫌疑，根据结果如何进行适当取舍，优化因子。

请大家指点：

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	1													
B	.851**	1												
C	.567**	.708**	1											
D	.545**	.627**	.700**	1										
E	.351*	.305	.588**	.652**	1									
F	.390*	.263	.403**	.569**	.746**	1								
G	.273	.238	.375*	.532**	.550**	.561**	1							
H	-.111	-.049	-.114	-.288	-.132	-.199	-.009	1						
I	-.205	-.218	-.235	-.329*	-.085	-.187	-.012	.857**	1					
J	.285	.292	.108	.129	-.229	-.089	.160	.148	-.097	1				
K	-.363*	-.144	-.126	-.267	-.274	-.280	-.091	.296	.095	.187	1			
L	-.302	-.195	-.384*	-.406**	-.513**	-.551**	-.396*	.295	.305	.046	.112	1		
M	-.056	.018	-.018	-.086	-.130	-.243	-.096	.426**	.450**	.165	-.081	.611**	1	
N	.001	-.001	-.175	-.256	-.302	-.179	-.015	.326*	.253	.432**	-.023	.280	.404**	1

**. 在 .01 水平（双侧）上显著相关。

*. 在 0.05 水平（双侧）上显著相关。

解 答 栏

楼主做这个的目的不外乎就是想判断这些指标是否存在严重多重共线性，如果存在，想进一步用因子分析/主成分分析降维。而要判定指标间是否存在严重多重共线性，用皮尔逊相关系数是不好判断的（变量间相关很正常，但我们不能通过变量相关就判定变量间存在严重多重共线性），要判定变量间是否存在严重多重共线性。楼主可以直接以一个自变量作为因变量，以其它自变量作为自变量，做一个回归的，回归时勾选共线性诊断对话框，然后看结果的 VIF 值的，大于 10 的表明存在多重共线性。然后用主成分分析对存在多重共线性的指标进行合并即可。祝好运~



提 问：

多个连续自变量对一个连续自变量的影响分析

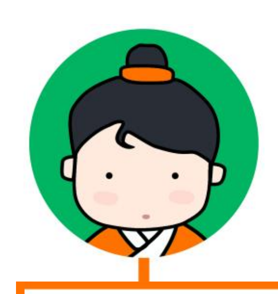
我现在想分析 5 个自变量对因变量的影响贡献程度，其中某几个自变量之间也相互影响。并且变量都是连续的，但没有时序性，尤其是因变量不是单纯的好或者坏，而是有不同数值表示的。我应该用什么分析呢，需要先用因子分析去除自变量之间的相互影响吗？

解 答 栏

用多元线性回归即可。不过在做回归时，需要对自变量进行多重共线性诊断。

追问：能消除多重共线性吗？

回答：还是回到了第一条回答，在构建模型时，对自变量间是否存在严重多重共线性进行诊断。如果不存在，直接建模就 OK；如果存在，用主成分回归或步进法即可。祝好运~



提 问：

如何用 STATA 命令消除多重共线性问题

大家好！我在做自变量之间的相关性分析时，发现资产负债率指标与其他变量之间存在相关系数大于 0.5 的情况，这是说明变量之间存在多重共线性吧，但是在不删除资产负债率这一指标的情况下，那么我该如何使用 STATA 对 LEV 处理，使其消除多重共线性啊？期待您的答复，谢谢。

解 答 栏

做变量间的相关只能大致判断变量间是否存在严重多重共线性，还是用专门的统计量检验为好。建议楼主用 estat vif 或者 coldiag2、collin 等命令去辅助判断下变量间是否真的存在严重多重共线性（完全多从共线性 STATA 建模时会直接 omit 掉）。若不存在，直接建模即可；若存在，可考虑用主成分回归，或者用逐步回归，或者用岭回归等方法处理。祝好运~



提 问：

共线性消除问题

gen CF = (ebitda - interest - tax - dividends)/ta

gen DPD = dividends!=.

我在数据库里面使用了这两行指令 然后这两个有 collinearity，但是 CF 的观察值只有 6000+， DPD 的观察值有 18000+

有没有什么办法可以消除掉这两个的 collinearity

解 答 栏

消除多重共线性的方法：一是删除其中一个变量，只将一个变量纳入模型；二是用主成分分析，将存在共线性的变量合并为一个新的变量，然后再纳入模型；三是用岭回归。祝好运~



提 问：

线性回归中的共线性诊断要怎么看

SAS 做了多元线性回归后的共线性诊断要怎么看。还有这里的三行 1、2、3 分别是什么含义？谢谢！

共线性诊断					
个数	特征值	条件指数	偏差比例 Intercept	price	temperature
1	2.93222	1.00000	0.00009813	0.00010075	0.01071
2	0.06734	6.59865	0.00197	0.00221	0.96571
3	0.00043315	82.27736	0.99793	0.99769	0.02358

解 答 栏

特征值的判定：特征值约为 0 表明可能存在多重共线性；条件指数的判定：条件指数大于 10 时表明可能存在多重共线性。所以综合而言，楼主应该注意下第 3 维度，其特征值接近于 0，且条件指数 80 多，远远高于 10。祝好运~



提 问：

关于多重共线性的问题

大家好。刚刚用SPSS做出来的一个结果 ,大家能帮忙看看吗？质量怎么样？这个共线性严重吗？可以接受吗？

系数[a]

模型		非标准化系数 B	非标准化系数 标准 误差	标准系数 试用版	t	Sig.	共线性统计量 容差	共线性统计量 VIF
1	(常量)	-4.669E-016	.028		.000	1.000		
	Zscore(VAR00002)	1.254	.163	1.254	7.689	.000	.029	35.057
	Zscore(VAR00003)	-.901	.151	-.901	-5.986	.000	.033	29.859
	Zscore(VAR00004)	.088	.036	.088	2.433	.015	.575	1.740
	Zscore(VAR00005)	.048	.031	.048	1.536	.125	.791	1.264
	Zscore(VAR00006)	.351	.052	.351	6.709	.000	.278	3.603
	Zscore(VAR00007)	.218	.073	.218	2.998	.003	.143	6.998
	Zscore(VAR00008)	-.057	.028	-.057	-2.028	.043	.946	1.057
	Zscore(VAR00009)	-.011	.033	-.011	-.348	.728	.716	1.396

a. 因变量: Zscore(VAR00001)

解 答 栏

Vif〉10 的两个变量存在严重多重共线性，需要处理。可删除一个或者用主成份回归解决。祝好运~



窥探计量与统计的方圆

六、

窥探计量与统计的方圆：回归分析精彩答疑 2 与资源荟萃

〖我也慢慢发现——计量学习虽重要，但相较而言，文章要解决的关键科学问题可能更重要。前者之所以重要，是因为它是解决关键科学问题的工具，而后者之所以重要，是因为它是文章的灵魂，可能是决定文章被录用与否的关键。有了这个心态的转变，我慢慢地开始反思这一路的计量自学之旅。〗

6.1

Likert 量表

6.1.1 概念

李克特量表（英语：Likert scale，在心理学上的读法相似于 'lick-urt' 与 'lie-kurt'之间，虽然前者的读法是取自于建立此量表的人的姓氏）是一种心理反应量表，常在问卷中使用，而且是目前调查研究（survey research）中使用最广泛的量表。当受测者回答此类问卷的项目时，他们具体的指出自己对该项陈述的认同程度。此量表是由 Rensis Likert 所建立。

李克特量表是使用各种李克特选项的总称。因为李克特选项，常常是一个视觉化量表（例如，在一个题目上的一条水平线，让受测者以画圈或点选的方式回答），这些选项有时也称为量表。但是，这容易造成混淆，因此，比较好的做法是，李克特选项专指一个单独的选项。

一个李克特选项是一个陈述。受测者被要求指出他或她们对该题目所陈述的认同程度，或任何形式的主观或客观评价。通常使用五个回应等级，但许多计量

心理学者（psychometrician）主张使用 7 或 9 个等级。5 等级、7 等级和 10 等级选项的数据，在简单的资料转换后，其平均数、变异数、偏态和峰度都很相似。例如李克特的五等选项的：

1.强烈反对

2.不同意

3.既不同意也不反对

4.同意

5.坚决同意

李克特量表是有两个极端的量化方法，衡量一个陈述的正面或负面回答。当中间选项“无意见”不能用时，有时会使用四等量表——一个强迫选择的方法。李克特量表也许会受到几种因素干扰而失真。受测者也许会回避勾选极端的选项（central tendency bias）；对陈述的习惯性认同(acquiescence bias)；或试着揣摩并迎合他们自己或他们的组织希望的结果（social desirability bias）。

回避勾选极端的选项（central tendency bias），特别是为了

回避被认为有极端主义者思想的可能 (比如下面将讨论的 social desirability bias 或者/和

为了回答简单，尤其是预计接下来的回答中会出现更强的观点，比如为了给测试中更强的回应而在早期回答中“留下空间”，这种预计会导致恶性偏差，很难

消除；

对陈述的 acquiescence bias，这种效应会在某些人中特别强，比如老弱病残中，他们存在一种鼓励并且想要取悦他人的制度文化。

否定某些带有防备趋向的问题，以防止回答错误或者消极后果，比如说答题者担心从自己的答案中找到自相矛盾的，特别是解释错误的或者语境偏离。

提供某些答案表示自己的强势或者并不缺乏某些功能（装作很好）;

提供某些答案表示自己的弱势或者缺乏某些功能（装作很差）;

试着揣摩并迎合他们自己或他们的组织希望的结果(social desirability bias 或者 norm defiance，上文提及的客观“装作很好”或者“装作很差”的主观表现)。

6.1.2 计量分析

在问卷完成后，每一个选项也许会被个别的分析，或某些成组的选项被加总并建立成一个量表。因此，李克特量表常常被称为累加量表(summative scale)。

至于个别的李克特选项可视为区间数据，或只应该被视为顺序数据，仍然是具争议性的议题。许多人将这样的项目视为顺序尺度的数据，因为特别是只有使用 5 个等级时，无法让受测者察觉到这些相邻的项目，其间隔是等距的。在另一方面，通常（正如上面的例子）其 response levels 的措辞清楚的暗示出中间类别的 response levels 的对称性；在最低限度，这样一个项目，将变成介于顺序

和区间尺度之间 ；只将它视为顺序数据将遗失一些信息。此外，如果该项目附带视觉近似评价标尺（visual analog scale），其回答程度的间隔则明确表示，其作为区间数据的论点是更加坚固。

当被视为顺序数据，李克特数据可以整理成长条图，以中位数或众数（但不是平均数）表现集中趋势，以四分位距表现分散程度（但不是标准差），或用非参数检验分析，如 Chi-square test, Mann-Whitney test，威尔克科逊检验，或 Kruskal-Wallis test。[3]

几个李克特题目的数据也许会被加总，若所有题目使用相同的李克特量表，则该量表可有效的接近区间尺度，此时可以将之视为区间数据测量潜在变项。如果加总结果满足这些假设，可以用参数统计（parametric statistical）如变异数分析作测试。但只有当项目在 5 个以上才可使用。

由李克特量表获得的数据，有时会合并所有的同意和不同意的回复为"接受"和"不接受"两个类别，此时会成为名目尺度。Chi-Square，Cochran Q，或 McNemar-Test 都是在资料做这些转换后常用的统计方法。

Consensus based assessment (CBA) 可以用来为没有普遍接受的标准或客观标准的领域产生一个客观的标准。CBA 可用于完善或甚至验证普遍接受的标准。

资料来源：

https://zh.wikipedia.org/wiki/%E6%9D%8E%E5%85%8B%E7%89%B9%E9%87%8F%E8%A1%A8

https://en.wikipedia.org/wiki/Likert_scale

http://bbs.pinggu.org/thread-447451-1-1.html

6.2

信度分析

信度分析

信度（Reliability）即可靠性，它是指采用同样的方法对同一对象重复测量时所得结果的一致性程度。信度指标多以相关系数表示，大致可分为三类：稳定系数（跨时间的一致性），等值系数（跨形式的一致性）和内在一致性系数（跨项目的一致性）。信度分析的方法主要有以下四种：重测信度法、复本信度法、折半信度法、α信度系数法。

重测信度法

这一方法是用同样的问卷对同一组被调查者间隔一定时间重复施测，计算两次施测结果的相关系数。显然，重测信度属于稳定系数。重测信度法特别适用于事实式问卷，如性别、出生年月等在两次施测中不应有任何差异，大多数被调查者的兴趣、爱好、习惯等在短时间内也不会有十分明显的变化。如果没有突发事件导致被调查者的态度、意见突变，这种方法也适用于态度、意见式问卷。由于

重测信度法需要对同一样本试测两次，被调查者容易受到各种事件、活动和他人的影响，而且间隔时间长短也有一定限制，因此在实施中有一定困难。

复本信度法

复本信度法是让同一组被调查者一次填答两份问卷复本，计算两个复本的相关系数。复本信度属于等值系数。复本信度法要求两个复本除表述方式不同外，在内容、格式、难度和对应题项的提问方向等方面要完全一致，而在实际调查中，很难使调查问卷达到这种要求，因此采用这种方法者较少。

折半信度法

折半信度法是将调查项目分为两半，计算两半得分的相关系数，进而估计整个量表的信度。折半信度属于内在一致性系数，测量的是两半题项得分间的一致性。这种方法一般不适用于事实式问卷（如年龄与性别无法相比），常用于态度、意见式问卷的信度分析。在问卷调查中，态度测量最常见的形式是 5 级李克特（Likert）量表。进行折半信度分析时，如果量表中含有反意题项，应先将反意题项的得分作逆向处理，以保证各题项得分方向的一致性，然后将全部题项按奇偶或前后分为尽可能相等的两半，计算二者的相关系数（即半个量表的信度系数），最后用斯皮尔曼-布朗（Spearman-Brown）公式：

$$\rho_{xx'}^{*} = \frac{N\rho_{xx'}}{1 + (N-1)\rho_{xx'}}$$

求出整个量表的信度系数。

α信度系数法

Cronbach's alpha 信度系数是目前最常用的信度系数，其公式为：

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2}\right)$$

其中，K 为量表中题项的总数，$\sigma_{Y_i}^2$为第 i 题得分的题内方差，σ_X^2为全部题项总得分的方差。从公式中可以看出，α系数评价的是量表中各题项得分间的一致性，属于内在一致性系数。这种方法适用于态度、意见式问卷（量表）的信度分析。

Cronbach's alpha	内部一致信度
α ≥ 0.9	Excellent
0.9 > α ≥ 0.8	Good
0.8 > α ≥ 0.7	Acceptable
0.7 > α ≥ 0.6	Questionable
0.6 > α ≥ 0.5	Poor
0.5 < α	Unacceptable

6.3

重要性-绩效分析(Importance & performance Analysis)

重要性-绩效分析(Importance & performance Analysis)是分析顾客对供方的产品和服务的重要性和绩效的感知，从而找到提高顾客满意度和顾客忠诚度的途径。

6.3.1 重要性-绩效分析的程序

该分析的程序如下：

1.第一步，数据收集。

首先是利用头脑风暴法等方法确定你的顾客、产品和服务。然后选择重要顾客，并请他们给产品或服务的好坏(即绩效)打分，分值范围从 1～5 分。同时，让顾客给产品或服务对他们的重要性打分，分值范围从 1～5 分。评分的具体标准如下表 1：

绩效	分值	重要性	分值
产品和服务足够好	3分	有一些重要	3分
有待提高	2分	愿意拥有，但不必需	2分
缺乏一致性	1分	不需要	1分

表1:产品及服务分值表

第二步，以同样的标准，自己(供方)给产品和服务的重要性及绩效打分。并将顾客的打分和你自己的打分记入见表 2

服务	供方给重要性打分	顾客给重要性打分	供方给绩效打分	顾客给绩效打分

表2:重要性-绩效评分表

第三步，利用矩阵进行重要性——绩效分析。

(1)顾客的评价分析，如图 1 所示。

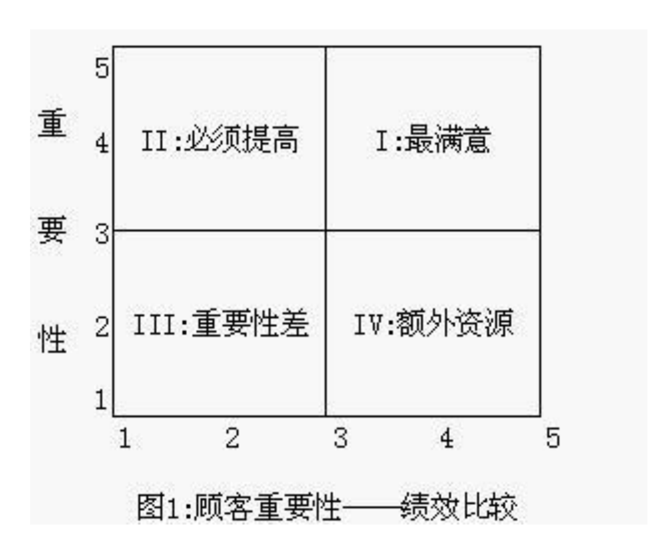


图1:顾客重要性——绩效比较

位于象限 I 表示最令人满意。顾客认为位于这一象限的产品或服务对他们最重要，而且做得很不错。

位于象限 II 表示需待改进。顾客认为位于这一象限的产品或服务对他们重要，但做得不够好。

位于象限 III 中的产品或服务需要改进，但重要性程度低于象限 II 中的产品或服务。

位于象限 IV 中的该产品或服务的资源可以转移到其他更重要的产品和服务中去，因为这一象限的产品或服务做得好，但重要性不高,使之做得好的资源可以转移出去。

(2)顾客评价的重要性与自己评价的重要性比较分析，如图 2 所示

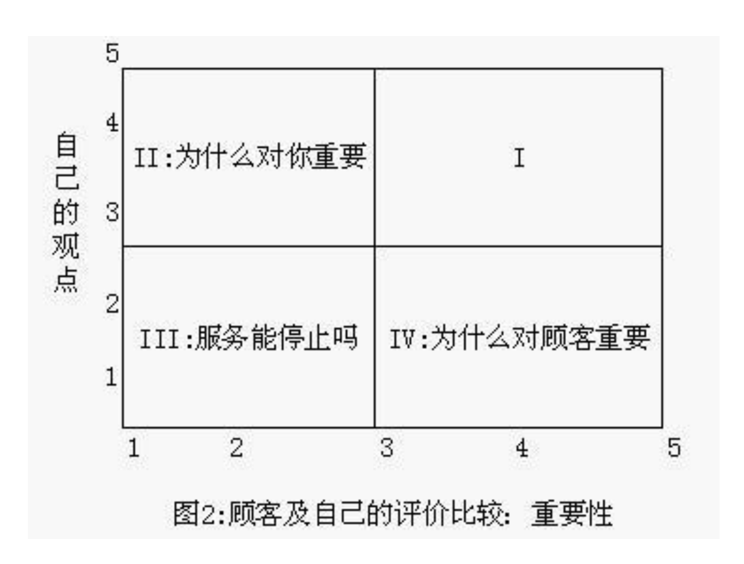


图2:顾客及自己的评价比较：重要性

位于象限 II 和 IV 中的产品和服务，顾客的评价和你的评价不一致，你与顾客有很大的沟通空间；

位于象限 III 中的产品和服务，顾客和你的重要性打分都是 1 或 2，你应该与顾客讨论用更重要的产品或服务代替现在的产品或服务。

(3)顾客评价的绩效与你评价的绩效分析，如图 3 所示。

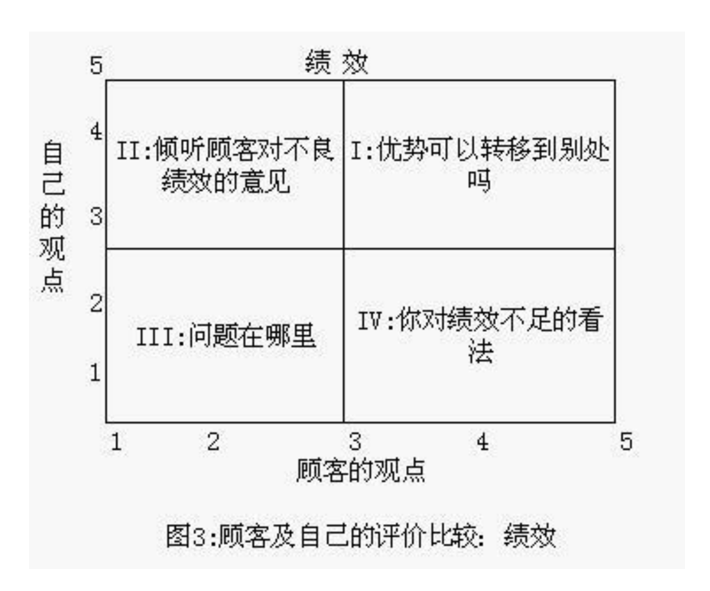


图3:顾客及自己的评价比较：绩效

位于象限 II 和 IV 的不一致说明与顾客还有很大的沟通空间。

位于象限 III 的产品或服务是你和顾客都认为绩效不足的，应该找到问题所在。

位于象限 I 的产品或服务是你和顾客都认为好的，也应该找出原因，从中发现别的服务可以借鉴的经验。

适用范围

这是一个与顾客互动的工具，它可以用于：

(1)分析顾客满意度。

(2)从顾客处获取关于产品和服务改进的有用信息。

(3)寻找顾客不满意的原因。

(4)做到合理配置服务资源。

(5)制定客户服务策略。

资料来源：

http://wiki.mbalib.com/wiki/%E9%87%8D%E8%A6%81%E6%80%A7-%E7%BB%A9%E6%95%88%E5%88%86%E6%9E%90



提 问：

关于李克特量表的问题求教。

现在要做一个李克特量表的文献综述，谁能给解释下五级、七级量表的差异，最好能提供点这方面的文献。先谢过大家了!

解 答 栏

没有关注过这个问题，但据我个人理解：里克特 5 级量表和 7 级量表差别不大，无非是 7 级量表测量的更精确些，但是正因为其更精确，所以测量起来会更困难。如果楼主非要去区分，那建议从测量精确性和成本等角度去考虑这个问题。PS:我所知道的关于量表的东西来源于一本书，推荐给楼主。吴明隆老师《问卷统计分析实务——SPSS 操作与应用》，链接：http://bbs.pinggu.org/thread-3017914-1-1.html。祝好运~



提 问：

问卷量表怎么进行相关分析

请问大神们，问卷的五级李克量表怎么进行相关分析？比如衡量 IT 能力的有 5 个题目，发展前途有 5 个题目，工资有 5 个题目。怎么进行这三个变量的相关分析?

解 答 栏

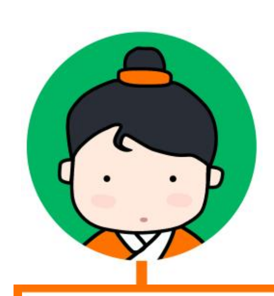
建议用主成份回归。即先对 15 个变量做信度检验，然后做主成分分析，提取出主成份后再进一步用公因子做相关分析或回归分析。

追 问：

谢谢你的解答，再问哈你就是如果我要做 8 个维度的相关分析，8 个维度分别有几个小题。是提取出几个因子后做因子的相关分析？ 因子的值是用旋转后成分矩阵的值乘以对应的变量得到的吗？ 还是说用因子的得分做相关分析，SPSS 里可以直接得到因子得分

回　答：

建议你用因子综合得分，通过主成分分析得到的这个因子得分，再经过一步计算得到最后的因子综合得分，然后再做相关分析。各个维度的综合得分=主成分分析得到的因子得分*该公因子方差占总方差比例。此处若做了因子旋转，用旋转后的结果计算即可，前后对应。祝好运~



提 问：

身高体重超过问卷限定值怎么统计？

问卷收集了身高和体重指标。比方说身高“124cm 及以下，125,126，……，190,191cm 以上”，体重“24KG 一下，25,26……90,91KG 及以上”。本来想根据身高和体重数据计算一些身体指标(BMI),但是有些填写的身高是“124cm 及以下”，体重有的填写“91KG 及以上”，填写这些的人就没办法算他们的 BMI，因为没有具体的数据。这种情况该怎么处理呢？

解 答 栏

这种数据有可能是无效数据，建议删除。

追　问：

谢谢回复。请问你是觉得身高填写的太极端了么？因为我有收集未成年人的数据，所以也有可能的。关键是这种数据问题该怎么解决~~

回　答：

正如楼主所说，由于没有临界值以下或以上的数据，所以没法根据公式对一些个体的 BMI 进行计算。出现这种情况，我个人认为的解决方法有二：一是你在做的事情中交代你关注的个体是区间以内的个体（如身高上线下线分别为 124cm 和 192cm，那么你就直接说你只关注 124-192cm 间的个体。其它的不关注）；二是求上下线的临界点（下线 124cm 对应的 BMI 临界点为 a，上线 192cm 对应的 BMI 临界点为 b)，然后将没有数据对应的上下线个体的 BMI 分别标记为低于 a 和高于 b。然后再进一步处理。祝好运~



提 问：

因变量是固定成绩，如何与李克特五分法取得的变量做相关呢？

自变量 F1 与 F2 是来自教师感知的变量（样本 235），因变量 Y 是来自学生的成绩（五所样本学校近三年的成绩），能否做相关？？

如何处理因变量使得适合做各种相关分析呢？

解 答 栏

看楼主具体想做什么。如果将学生成绩看做是连续变量，用多元线性回归即可；如果将学生成绩按照一定的标准分类（如分为高中低三等），那用有序多分类 Logistic 回归即可。

追 问:

这个学生成绩不是单个人的线性上升数据，而是全体的某次成绩，所以尝试将五个学校抽样分别取三年前 50 名的平均值，构成一个 250 的样本，在做回归分析的时候，与其余两个变量 N 不一样，也能做分析吗？

另外这些数据是否需要打乱？因为每个学校累加，就都是五个层面的从高到低，SPSS 也没有随机分布数据这一功能呢

这些变量不会被看做是一一对应的变量的吗？

因为方便抽样，在每所学校也都抽取了前五十名的学生，一开始想到学校一一对应教师问卷，但是录入时已经打乱了所有学校问卷，只能做整体的研究。

目前比较惆怅是，F1 与 F2 已经证明具有相关性，无法避免多元共线性问题了。是不是就得该做最简单的回归分析了？

我想做 F1 或 F2 对因变量有没影响，如果 F2 为 F1 的调节变量或中介变量，这又需要用什么分析证明？

回　答:

里面问题太多，认真梳理了一下。个人的建议如下：

第一个问题，在做回归分析时，自变量和因变量样本不一样。你的自变量是对应于教师感知的变量 F1 与 F2，因变量是抽样的学生成绩。这种数据类型有点类似于嵌套型的数据了，即学生为个体，教师与班级层次对应，一个教师可能有很多学生，这种情况下即使自变量与因变量样本不等，也可以用层次线性模型进行分析。

第二个问题，数据是否需要被打乱这些变量不会被看做是一一对应的变量的吗？我个人觉得打乱不打乱都没多大影响。同时，对应于第一个问题的回答，教

师和学生可能是一对多的关系，建议用层次线性模型进行分析。

第三个问题，F1 与 F2 已经证明具有相关性，无法避免多元共线性问题了。是不是就得该做最简单的回归分析了？变量之间存在相关性很正常，只要不造成严重多重共线性就可以，建议在后续建模分析时对 F1 与 F2 进行多重共线性诊断，然后再进行下一步分析。

第四个问题，我想做 F1 或 F2 对因变量有没影响，如果 F2 为 F1 的调节变量或中介变量，这又需要用什么分析证明？这涉及到调节效应和中介效应的问题，三言两语说不清楚。推荐谢宇老师《回归分析》（链接：http://bbs.pinggu.org/thread-3065280-1-1.html）。祝好运~



提 问：

信度分析克朗巴哈 a 系数为 0.7，不知这个信度系数是否可以？

今天做信度分析，克朗巴哈 a 系数为 0.7，不知这个信度系数是否可以？

解 答 栏

通常 Cronbach α系数的值在 0 和 1 之间。如果α系数不超过 0.6，一般认为内部一致信度不足；达到 0.7-0.8 时表示量表具有相当的信度，达 0.8-0.9 时说明量表信度非常好。Cronbach α系数的一个重要特性是它们值会随着量表项目的增加而增加，因此，Cronbach α系数可能由于量表中包含多余的测量项目而被人为地、不适当地提高。还有一种可以和 Cronbach α系数同时使用的系数。系数能够帮助评价，在计算 Cronbach α系数的过程中，平均数的计算是否掩盖了某些不相关的测量项目。不同的研究者对信度系数的界限值有不同的看法，有学者认为，在基础研究中 Cronbach α系数至少应达到 0.8 才能接受，在探索研究中 Cronbach α系数至少应达到 0.7 才能接受，而在实务研究中，Cronbach α系数只需达到 0.6 即可。祝好运~



提 问：

SPSS 在做社会调研时，怎样分析问卷有效性。

在做社会问卷调查时，怎样检验数据的真实性，这样做的理由及作用

解 答 栏

我个人主要是做社会调查研究的，以下是个人关于这方面的一点心得，供参考。第一，如果社会调查是以量表的形式测度的，那么可以用信度检验去检验数据的一致性；第二，如果数据不是量表形式的，那么看看设计的问卷里面是否前后有内容相关的选项可以相互验证（如我是做农户调研的，我前面问了一个问题是农户是否掌握有技能，被访问农户开始说没有。后来我问他干什么工作，他说厨师。其实他前面的回答就是有问题的。他可能没有理解我前面问是否掌握有技能这个问题，所以再倒回去核实一下就能确定最终答案了）。在我的印象中，一般的问卷设计前后的问题多多少少都会有些关联。你可以以此辅助判断一下；第

三，对数据做描述性统计分析，然后看看里面的极端异常值，然后再结合问卷去检验一下这些值大概是怎么来的，是调研员笔误还是其它原因。一般而言，大多数数据都会在均值左右波动，如果有很多数据波动太大，可能就是有问题的；第四，如果真有人根据常识仿造数据，那可能是查不出来的。这个只能看个人的学术道德了。祝好运~



提 问：

求助！一个调查问卷中的统计分析问题，如何对比讲解前后人们观点的变化？

想通过问卷调查对比一个讲解工作前后，人们对同一问题的看法是否有改变。例如其中一个问题是："您认为在这个过程中，动物是否展示出了自然行为？" 1.听讲解前______2. 听讲解后________（备选项：A 我不知道；B 没有展现；C 展现了一些；D 展现了很多）。

每一个参与调查的游客都填写了参与活动前后的看法，所以数据可以说是

配对的。因为我在统计上是小白，现在想请问各位大神们，这种情况下，应该采用什么统计方法，对比这种前后选项的差异性呢？用卡方检验可以么？非常感谢大家的帮助！！

解 答 栏

由于变量类型是类别变量，所以用卡方检验就好。建议楼主看看张文彤老师《SPSS 统计分析教程》（链接：http://bbs.pinggu.org/thread-3573256-1-1.html），上面有原理介绍，有操作截图，有结果解释。祝好运~



提 问：

统计分析中如何处理打分题目的分析

在问卷中，例如：对于你在找工作过程中，请将您得到的信息来源，按重要程度打分，满分为 5 分，依次递减：数据录入中，将每一个陈述设为一个变量，1-5 分设置成不同选项，就想知道这 8 个陈述中，他们的重要性如何排列，请问怎样处理呢？

解 答 栏

单独的比较某两个词条（如在社会实践中获得的机会与亲友等私人关系推荐介绍）没多大意义，因为如果同样打的都是 5 分，没法比较。现在可以比较的是调研样本个体在各个词条上的均值，以此差异来比较各个词条的重要性。如有 100 个人参与了此项问卷调查。在社会实践中获得的机会的得分均值为 4.2，而亲友等私人关系推荐介绍得分为 4.8，那么可粗略认为后者比前者更重要。祝好运~



提 问：

关于量表设计的“降维”

请问，1.量表设计过程中的“降维”是指减少题项的数目还是减少因子的数目啊？

2.因子是不是就是构念?

解 答 栏

问题 1：量表设计中的降维一般指量表词条经过信度检验后用因子分析对存在严重多重共线性的量表词条进行“合并”的过程，其实就是指做因子分析；问题 2：对量表词条做因子分析后，得到公因子，根据特征对其命名后，如果要进一步构建结构方程模型，那么得到的公因子可称之为构念。祝好运~



提 问：

指标体系里部分指标数据通过统计获取，部分指标数据发问卷收集，想求各指标权重，请问什么方法能够做到？

指标体系里部分指标数据通过统计获取，部分指标数据发问卷收集，想求各指标权重，请问什么方法能够做到？

解 答 栏

求指标权重方法比较多，分为主观权重法和客观权重法。前者如层次分析法；后者有因子分析法、熵值法等。祝好运。

http://bbs.pinggu.org/forum.php?mod=viewthread&tid=3809006，可参考下福荣山版主对各类求权重方法的归纳。祝好运~



提 问：

问卷调查针对自变量、中间变量和因变量，问卷发放对象各为一类，对实证结果科学性有何影响

问卷调查针对自变量、中间变量和因变量，问卷发放对象各为一类，对实证结果科学性有何影响，在不更换调查对象的前提下，如何解决？

解 答 栏

对象不一样，结果肯定会受很大的影响。举个典型例子：测两个时间序列数据，一个是人的身高，一个是树的高度。用人的身高和树的高度做相关分析/回归分析，结果发现这二者高度相关。于是得到结论：人的身高随着树的高度变化而变化，这显然是有问题的（因为这二者间其实是没有关系的）。所以套在你这个问题上，基本也是一个理。祝好运~

6.3.2 相关分析

相关分析是研究现象之间是否存在某种依存关系，并对具体有依存关系的现象探讨其相关方向以及相关程度，是研究随机变量之间的相关关系的一种统计方法。

社会经济现象之间存在着大量的相互联系、相互依赖、相互制约的数量关系。这种关系可分为两种类型。**一类**是函数关系，它反映着现象之间严格的依存关系，也称确定性的依存关系。在这种关系中，对于变量的每一个数值，都有一个或几个确定的值与之对应。**另一类**为相关关系，在这种关系中，变量之间存在着不确定、不严格的依存关系，对于变量的某个数值，可以有另一变量的若干数值与之相对应，这若干个数值围绕着它们的平均数呈现出有规律的波动。例如，批量生产的某产品产量与相对应的单位产品成本，某些商品价格的升降与消费者需求的变化，就存在着这样的相关关系。

- **相关分析的分类：**

1、线性相关分析：研究两个变量间线性关系的程度。用相关系数 r 来描述。

（1）正相关：如果 x,y 变化的方向一致，如身高与体重的关系，$r>0$；一般地，

·$|r|>0.95$ 存在显著性相关；

·$|r|\geq0.8$ 高度相关；

·$0.5 \le |r| < 0.8$ 中度相关；

·$0.3 \le |r| < 0.5$ 低度相关；

·$|r| < 0.3$ 关系极弱，认为不相关

（2）负相关：如果 x,y 变化的方向相反，如吸烟与肺功能的关系，$r<0$；

（3）无线性相关：$r=0$。

如果变量 Y 与 X 间是函数关系，则 $r=1$ 或 $r=-1$；如果变量 Y 与 X 间是统计关系，则 $-1<r<1$。

（4）r 的计算有三种：

①Pearson 相关系数：对定距连续变量的数据进行计算。

②Spearman 和 Kendall 相关系数：对分类变量的数据或变量值的分布明显非正态或分布不明时，计算时先对离散数据进行排序或对定距变量值排（求）秩。

2、偏相关分析：研究两个变量之间的线性相关关系时，控制可能对其产生影响的变量。如控制年龄和工作经验的影响，估计工资收入与受教育水平之间的相关关系。

相关分析在统计软件中的操作也非常简单，以 SPSS 和 Eviews 举例来说：

- **SPSS 相关分析图例：**

分析-相关-双变量，打开双变量相关对话框,将 X 和 Y 选中导入到变量窗口，然后相关系数选择 Pearson 相关系数,也可以选择其他两个，点击确定在结果输

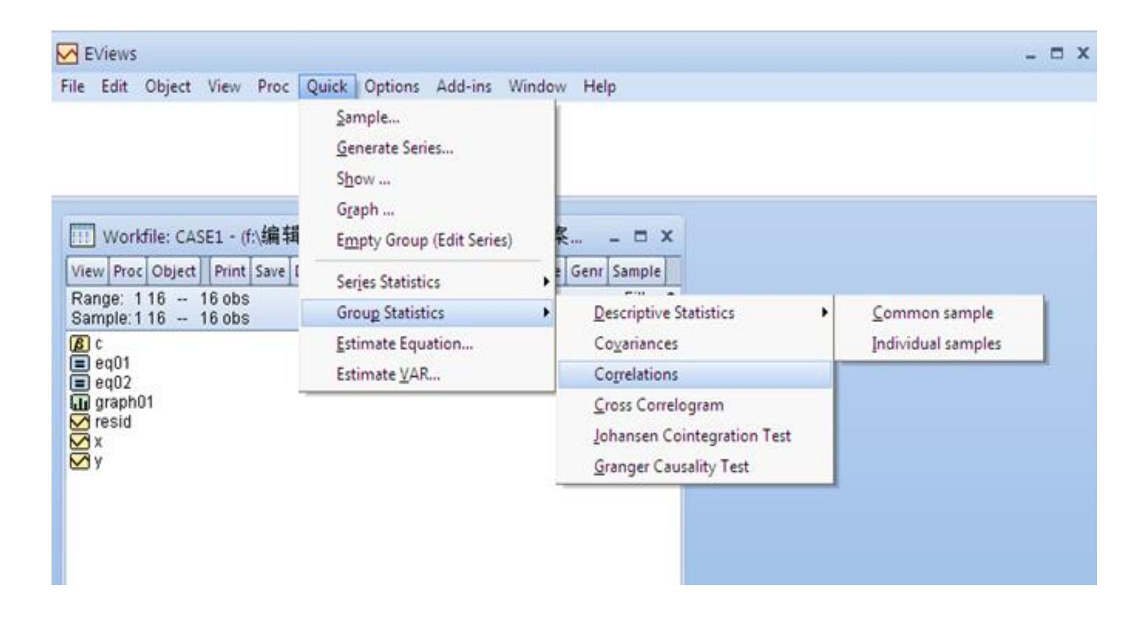
出窗口显示相关性分析结果。





而在 Eviews 中，操作也很方便：

通过 quick-group statistics-correlatins 即可得到相应的结果：





提 问：

SPSS 相关性分析

要对一份调查问卷的结果进行相关性分析，该怎么做呢？一点眉目都没用额，求帮助。

解 答 栏

用 SPSS“分析”——“相关”命令即可。方法的选择要具体看变量的类型：若变量是连续型的，用 Pearson 相关系数；若变量是类别变量或有序变量，用 Kendall 的 tau-b 或 Spearman 系数。推荐张文彤老师《SPSS 统计分析教程》，上面有详细原理介绍和截图操作展示，还有结果解读。链接如下：http://bbs.pinggu.org/thread-3573256-1-1.html。祝好运~



提 问：

求助 Pearson 相关系数和 Spearman 相关系数 STATA 命令分别是什么？

看了论坛里大家广泛讨论 pwcorr_a 命令来衡量变量之间的偏相关系数，但这个命令具体依据的是什么方法呢，？计量软件中 Pearson 相关系数和 Spearman 相关系数 STATA 命令分别是什么呢？还有结果的导出？初学者，恳请大神不吝赐教！!!

解 答 栏

Pearson 相关系数的命令为：correlate y x1 x2；Spearman 相关系数的命令为：spearman y x1 x2。祝好运~



提 问：

简单相关分析 三种检验方法得到的结果不一样

以下是我随便做的一组数据进行相关分析得到的结果：

相关性

		开盘价	涨幅
开盘价	Pearson 相关性	1	-.537
	显著性（双侧）		.059
	N	13	13
涨幅	Pearson 相关性	-.537	1
	显著性（双侧）	.059	
	N	13	13

相关系数

			开盘价	涨幅
Kendall 的 tau_b	开盘价	相关系数	1.000	-.410
		Sig.（双侧）	.	.051
		N	13	13
	涨幅	相关系数	-.410	1.000
		Sig.（双侧）	.051	.
		N	13	13
Spearman 的 rho	开盘价	相关系数	1.000	-.571*
		Sig.（双侧）	.	.041
		N	13	13
	涨幅	相关系数	-.571*	1.000
		Sig.（双侧）	.041	.
		N	13	13

pearson 相关性大于 0.5 说明显著相关 但是显著性却说明 不相关 应该怎么解读？

kendall 也同样。pearson 和 kendall 的显著性都大于 0.05 接受原假设 两组数据不相关 但是 spearman 的显著性却低于 0.05 说明两组数据相关我应该如何解读？求指点！

解 答 栏

表格中的三种方法对应于不同的数据类型的。楼主的指标（开盘价和涨幅）都是连续性变量，要看这两个变量间的相关直接看 pearson 相关系数这个表就行。

追 问:

首先谢谢！然后我还想问问，书中原例是用标准普尔和道琼斯 10 周内的收盘价做对比同时用了 3 种相关系数做检验 ,所以我才会拿和例子相似的股票的数据做练习。能说明一下为什么书中原例能用后两种方法做检验尔我数据的不行吗？

另外能帮我解读一下我的第一个表格所表明的含义吗？

感激不尽！

回　答:

pearson 相关系数是做连续性变量相关用的，Kendall 与 Spearman 相关系数是有序变量或类别变量相关用的。你的变量是连续性的变量，所以看 pearson 相关系数对应到表格即可。至于楼主说的其它有人用了这三个系数，在我个人所知道的统计/计量中，这么做是有问题的。至于第一个表的解读：开盘价和涨幅间存在负向相关（相关系数为-0.537），且在 0.1 水平上显著。祝好运~

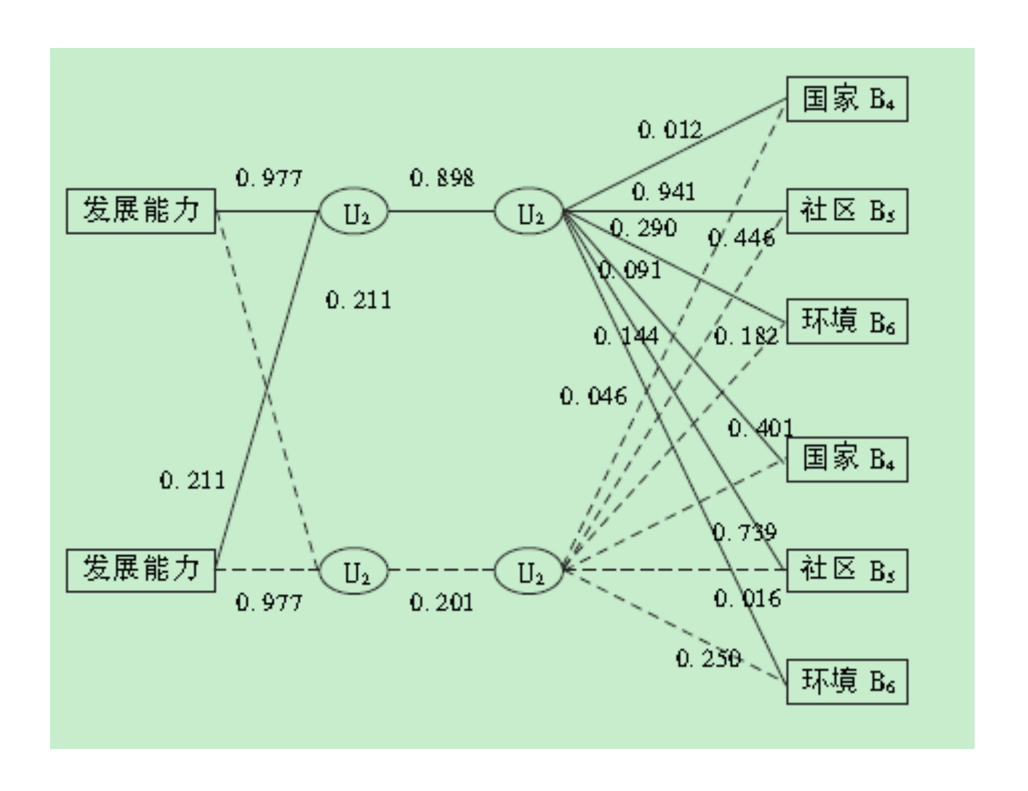


提 问：

典型相关分析的结构图怎么画

请问这个图是用哪个软件画的

求具体步骤和命令



解 答 栏

Amos 软件应该能做。推荐吴明隆老师《结构方程模型——Amos 的操作与应用》一书，里面有详细的原理和截图操作介绍，楼主按照操作来就能得到想要的。链接如下：http://bbs.pinggu.org/thread-1040309-1-1.html。祝好运~



提 问：

SPSS 中年龄的相关分析

实证研究中，建立结构方程模型，将年龄、性别、教育程度、收入水平以及在网站上的注册时间作为调查问卷对象的特征变量，现要将这些特征变量与其他变量进行相关分析，其他变量采用的李克特五分量表测量，请问如何使用 SPSS 处理？？？

解 答 栏

李克特五分量表可以近似看做连续性的变量处理，楼主想做的相关分析可以通过以下步骤实现："分析"——"相关"——"双变量"，将你想做相关分析的变量选到相应的对话框点击确认就行了。具体操作可参见张文彤老师《SPSS 统计分析教程》，链接如下：

http://bbs.pinggu.org/thread-3573256-1-1.html。祝好运~

6.3.3 因素方差分析

方差分析的基本原理是认为不同处理组的均数间的差别基本来源有两个：

1、 随机误差，如测量误差造成的差异或个体间的差异，称为组内差异，用变量在各组的均值与该组内变量值之偏差平方和的总和表示。

2、 实验条件，即不同的处理造成的差异，称为组间差异。用变量在各组的均值与总均值之偏差平方和表示。

方差分析的基本思想是 通过分析研究不同来源的变异对总变异的贡献大小，从而确定可控因素对研究结果影响力的大小。

如果用均方（离差平方和除以自由度）代替离差平方和以消除各组样本数不同的影响，则方差分析就是用组间均方去除组内均方的商（即 F 值）与 1 相比较，若 F 值接近 1，则说明各组均值间的差异没有统计学意义，若 F 值远大于 1，则说明各组均值间的差异有统计学意义。实际应用中检验假设成立条件下 F 值大于特定值的概率可通过查阅 F 界值表（方差分析用）获得。

- **方差分析主要用途：**

1、均数差别的显著性检验

2、分离各有关因素并估计其对总变异的作用

3、分析因素间的交互作用

4、方差齐性检验。

- **根据资料设计类型的不同，有以下两种方差分析的方法：**

1、对成组设计的多个样本均值比较，应采用完全随机设计的方差分析，即单因素方差分析。

2、对随机区组设计的多个样本均值比较，应采用配伍组设计的方差分析，即两因素方差分析。

- **两类方差分析的异同：**

两类方差分析的基本步骤相同，只是变异的分解方式不同，对成组设计的资料，总变异分解为组内变异和组间变异（随机误差），即：SS 总=SS 组间+SS 组内。

而对配伍组设计的资料，总变异除了分解为处理组变异和随机误差外还包括配伍组变异，即：SS 总=SS 处理+SS 配伍+SS 误差。

- **方差分析的基本步骤：**

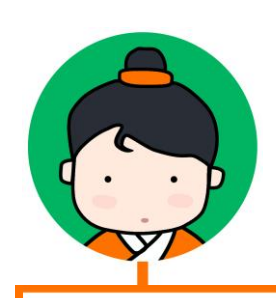
1、建立检验假设；

H0：多个样本总体均值相等；

H1：多个样本总体均值不相等或不全等。（一般选取检验水准为 0.05）。

2、计算检验统计量 F 值；

3、确定 P 值并作出推断结果。



提 问：

把好几种分类型数据对一个总体的影响具体表现出来？

怎么样把好几种分类型数据对一个总体的影响具体表现出来？

解 答 栏

没有太明白楼主意思。楼主可以介绍的详细一点（如因变量和自变量分别是什么类型的数据，楼主想做什么），然后我们才能给出更具体的建议（用什么方法，如何实现，结果如何解释等）。

追 问:

举个例子，就像调查饮食，生活习惯，价值观念等因素中男女差异的程度类似的

回 答:

若只是比较男女间的饮食、生活习惯、价值观念等差异，直接用 t 检验/方差分析/卡方检验就好。其中，连续性的变量组间差异比较用 t 检验/方差分析，类别变量间的差异比较用卡方检验就好。具体的原理和操作截图以及结果解释介绍推荐张文彤老师的《SPSS 统计分析教程》，链接：

http://bbs.pinggu.org/thread-3573256-1-1.html。祝好运~



提 问：

单因素方差分析方差不齐时该怎么办

单因素方差分析方差不齐，这个时候下面的单因素方差 p 值小于 0.05 还有效吗？是不是不能算有显著性差异了？

解 答 栏

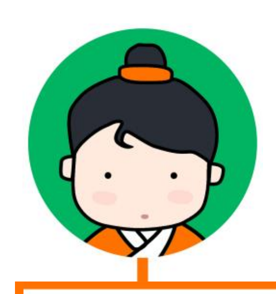
方差不齐时有相应的方法，SPSS 有 Tamhane’s T2 法勾选框，专门针对方差不齐时做方差分析用的。

追　问：

你好，我选了 Tamhane’s T2 法勾选框，那这样最终 SPSS 输出的结果 p<.05 就说明是有显著性差异的，然后看 Tamhane’s T2 一栏是哪组之间产生了差异，是这样吗？

回　答：

勾选 Tamhane’s T2 法后，出来的结果如何看，和平时方差齐性时做方差分析一样的，组与组之间比较看 p 值。祝好运~



提 问：

求教多水平分析

求问分析图书的销量影响因素，包括图书本身层面，作者层面，出版社层面。需不需要使用多水平分析，如果用，一本图书既属于作者，也属于出版社，可是作者和出版社又不是并列关系，而是从属关系。那么这种情况该如何做多水平分析。另外，因子和协变量的区别在哪啊。一个名义变量作为因子和作为协变量，结果有很大不同。

解 答 栏

作者将影响图书的销量因素分为三大类，分别为图书本身层面，作者层面，出版社层面。这种条件下直接用分层回归即可。说得简单点就是构建三个模型，第一个模型包含图书本身层面指标和图书的销量；第二个模型包含图书本身层面指标、作者层面的指标和图书的销量；第三个模型包含图书本身层面，作者层面，

出版社层面的指标和图书的销量。这是常见的做法。

还有种方法是用分层线性模型，这是基于书是第一层，属于某个作者；作者是第二层，又属于某个出版社。这种方法要复杂的多。

http://bbs.pinggu.org/thread-3035963-1-1.html，可看看张雷老师这本书，用 HLM 软件实现。

因子和协变量的区别在于：协变量框一般要求纳入连续性的变量，而因子框没这个限制，类别变量和连续性变量均可纳入；至于你说的一个名义变量作为因子和作为协变量，结果有很大不同。这是自然，因为在构建回归模型时，如果要纳入名义变量，要先将其处理为哑变量/虚拟变量才能纳入模型的。直接纳入模型是有问题的。祝好运~

6.3.4 非参数检验

非参数统计是统计学的一个重要分支，它在实践中有着广泛的应用。所谓统计推断，就是由样本观察值去了解总体，它是统计学的基本任务之一。若根据经验或某种理论我们能在推断之前就对总体作一些假设，则这些假设无疑有助于提高统计推断的效率。这种情况下的统计方法称为“参数统计”。如果我们所知很少，以致于在推断之前不能对总体作任何假设，或仅能作一些非常一般性(例如连续分布、对称分布等)的假设，这时如果仍然使用参数统计方法，其统计推断的结果显然是不可信的，甚至有可能是错的。在对总体的分布不作假设或仅作非常一般性假设条件下的统计方法称为“非参数统计”。

由于非参数统计方法与总体究竟是什么分布几乎没有什么关系，所以它的应用范围很广，它在社会学、医学、生物学、心理学、教育学等领域都有着广泛的应用。由于有关于总体的假设，所以参数统计的推断方法是针对这个假设的。相对而言，非参数统计的推断方法是很一般的，它仅应用样本观察值中一些非常直观（例如次序）的信息。所以非参数统计分析含有丰富的统计思想。

非参数统计最常用于具备下述特征的情况：

1、待分析数据不满足参数检验所要求的假定，因而无法应用参数检验。例如，我们曾遇到过的非正态总体小样本，在 t-检验法也不适用时，作为替代方法，就可以采用非参数检验。

2、仅由一些等级构成的数据，不能应用参数检验。例如，消费者可能被问及对几种不同商标的饮料的喜欢程度，虽然，他们不能对每种商标都指定一个数字来表示他们对该商标的喜欢程度，却能将几种商标按喜欢的顺序分成等级。这种情形也宜采用非参数检验。

3、所提的问题中并不包含参数，也不能用参数检验。例如，我们想判断一个样本是否为随机样本，采用非参数检验法就是适当的。

4、当我们需要迅速得出结果时，也可以不用参数统计方法而用非参数统计方法来达到目的。一般说来，非参数统计方法所要求的计算与参数统计方法相比，完成起来既快且易。有些非参数统计方法的计算，就算对统计学知识不熟练的人，也能在收集数据时及时予以完成。

- **非参数统计的优缺点：**

非参数统计与传统的参数统计相比，有以下优点：

1、非参数统计方法要求的假定条件比较少，因而它的适用范围比较广泛。

2、多数非参数统计方法要求的运算比较简单，可以迅速完成计算取得结果，因而比较节约时间。

3、大多数非参数统计方法在直观上比较容易理解，不需要太多的数学基础知识和统计学知识。

4、大多数非参数统计方法可用来分析如象由等级构成的数据资料，而对计

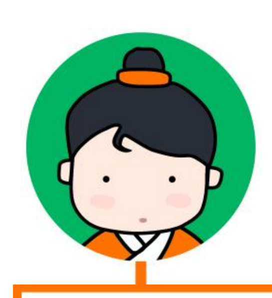
量水准较低的数据资料，参数统计方法却不适用。

5、当推论多达 3 个以上时，非参数统计方法尤具优越性。

非参数统计方法也有以下缺点：

1、由于方法简单，用的计量水准较低，因此，如果能与参数统计方法同时使用时，就不如参数统计方法敏感。若为追求简单而使用非参数统计方法，其检验功效就要差些。这就是说，在给定的显著性水平下进行检验时，非参数统计方法与参数统计方法相比，第Ⅱ类错误的概率β要大些。

2、对于大样本，如不采用适当的近似，计算可能变得十分复杂。



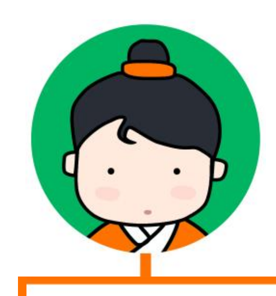
提 问：

求教，秩和检验用那个值比较好？

发表文章时写 u 值还是 Z 值合适一些？

解 答 栏

一般用 z 值比较多，楼主可结合相关专业文献来确定到底用哪个值。祝好运~



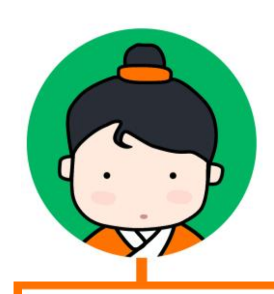
提 问：

非参数检验求问？

3 组不同干预方式处理环境物体表面，采样后进行细菌培养，得出 3 组菌落数。本人用 Kruskal-Wallis 检验分析 3 组间细菌菌落数之间是否存在差异，考虑到细菌是呈比例生长的，请问原始数据（细菌菌落数）是否要先经对数转换才能进行比较分析？

解 答 栏

如果用 t 检验或者方差分析需要考虑变量的分布，用非参数检验应该不用。不过感觉为了稳妥起见，楼主可以用箱图看看是否存在极端异常值，数据是不是右偏态，如果是，可考虑对其取对数。祝好运~



提 问：

特殊的独立样本 T 检验怎么做

在独立样本 T 检验中，检验变量都分组了，即用 1，2，3，4 等标志值表示，怎么去做这个呢，直接做的话，均值计算的是这个标志值

解 答 栏

楼主的方法可能用错了。连续性变量间的差异检验才用独立样本 t 检验，而你的关注变量是类别变量，需要用卡方检验检验其差异。推荐张文彤老师《SPSS 统计分析教程》，看看上面的相关部分的讲解吧。链接：

http://bbs.pinggu.org/thread-3573256-1-1.html。祝好运~

6.3.5 其他模型（灰色关联度、多层线性模型 HLM)

6.3.5.1 灰色关联度

灰色系统理论提出了灰色关联度的概念，它是指系统中两个因素关联性大小的量度，关联度的大小直接反映系统中的各因素对目标值的影响程度。运用灰色关联分析法进行因素分析的一般步骤为：

第一步：确定分析数列。

确定反映系统行为特征的参考数列和影响系统行为的比较数列。反映系统行为特征的数据序列，称为参考数列。（Y）设参考数列（又称母序列）为 Y={Y（k）|k=1，2，Λ，n}；影响系统行为的因素组成的数据序列，称比较数列。（X）比较数列（又称子序列）Xi={Xi（k）|k=1,2,Λ,n}，i=1，2，Λ，m。

第二步，变量的无量纲化

由于系统中各因素列中的数据可能因量纲不同，不便于比较或在比较时难以得到正确的结论。因此为了保证结果的可靠性，在进行灰色关联度分析时，一般都要进行数据的无量纲化处理。

$$x_i(k) = \frac{X_i(k)}{X_i(l)}, k = 1, 2, \Lambda, n; i = 0, 1, 2, \Lambda, m$$

第三步，计算关联系数。X0（k）与 xi（k）的关联系数

$$\xi_i(k) = \frac{\min_i \min_k |y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|}{|y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|}$$

记$\triangle_i(k) = |y(k) - x_i(k)|$，则

$$\xi_i(k) = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}$$

$\rho \in (0, \infty)$，称为分辨系数。ρ越小，分辨力越大，一般ρ的取值区间为(0,1)，具体取值可视情况而定。当$\rho \leq 0.5463$时，分辨力最好，通常取ρ=0.5。

ξi（k）继比较数列 xi 的第 k 个元素与参考数列 xo 的第 k 个元素之间的关联系数。

第四步，计算关联度

因为关联系数是比较数列与参考数列在各个时刻（即曲线中的各点）的关联程度值，所以它的数不止一个，而信息过于分散不便于进行整体性比较。因此有必要将各个时刻（即曲线中的各点）的关联系数集中为一个值，即求其平均值，作为比较数列与参考数列间关联程度的数量表示，关联度 ri 公式如下：

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \xi_i(k), k = 1, 2, \Lambda, n$$

第五步，关联度排序

关联度按大小排序，如果 r1<r2，则参考数列 y 与比较数列 x2 更相似。

在算出 Xi（k）序列与 Y（k）序列的关联系数后，计算各类关联系数的平均值，平均值 ri 就称为 Y（k）与 Xi（k）的关联度。

6.3.5.2 HLM 多层线性模型

在社会科学研究进行取样时，样本往往来自于不同的层级和单位，由此得到的数据带来了很多跨级（多层）。多层线性模型又叫做“多层分析（multilevel analysis）”或者是“分层线性模型（hierarchical liner modeling）”。

在社会科学中，多层线性的结构非常具有普遍性，如下图列出四种常见的情况



拿两层举例子，假如说现在我们考察学生自我效能感对学生成绩的影响，

所学校中抽取了 1000 名学生，那么很有可能的情况就是有些学校学生的自

我效能感平均值较高，而这就有可能是因为学校为贵族学校，学生的经济水平很高。而也可能有民工学校，经济水平较低，自我效能感普遍较低。那么这就存在一种情况就是学生的成绩受到学生个体的自我效能感影响，而每个学校的自我效能感可能与整个学校的整体经济水平有关。那么这就是学生嵌套在学校之间的例子。

多层线性模型的基本公式

第一层：

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + r_{ij} \quad (14)$$

第二层：

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \mu_{0j} \quad (15)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_{1j} + \mu_{1j} \quad (16)$$

拿上面的例子我们可以写出对于这个案例的多层线性模型。

第一层：学生成绩=β0+β1*学生自我效能感+r

第二层：β0=γ00+γ01*学校社会经济生活水平+μ1

β1=γ10+γ11*学校社会经济生活水平+μ2



提 问：

SPSS 多元回归分层回归的一些问题

小妹正准备开始写论文，用 SPSS 进行多元回归的分层回归分析，关于 SPSS 之前上课学过一些，完全是大白，分析过程中遇到以下问题，找了一些资料也不得而解，希望热心人士解答，不胜感激。PS:两个控制变量 A/B，三个自变量 C/D/E，一个因变量 F。

1.在分析前做了相关性分析，其中有一个控制变量 A 与一个自变量 C 不相关，其他各变量都是相关的，这个会影响分析结果吗？需要去掉这个控制变量吗？（控制变量的选择主要是参考了其他论文）

2.控制变量系数不显著表明什么？是不是代表应该去掉这个控制变量？

3.在第二层加入自变量时，F 值显著，自变量 C 系数 T 值不显著代表什么？C 和 F 明明相关，但为什么会不显著（后面要分析 E 的中介作用，中介作用貌似 C 和 F 显著是前提，那我该怎么办？把 C 去掉？可是没有其他变量代替 C 了）？

4.加入自变量后，控制变量系数从显著变成不显著了，这个要怎么解释，对结果分析有影响吗？

解 答 栏

第一个问题做自变量间的相关性分析是为了进一步判定自变量间是否存在多重共线性么？自变量间只要不存在完全或近似多重共线性，其有一定的相关或不相关都是很正常的，不用太纠结。ps：自变量间是否存在严重的多重共线性建议利用 vif 进行诊断（vif>10 一般认为变量间存在严重多重共线性，需要进一步处理），不要单看相关系数。

第二个问题，如果你的变量选取（包括控制变量）是基于理论或文献来，那即使控制变量不显著，也可以纳入模型。不过由于是控制变量，且它是外生变量（删除它不会对模型结果造成影响），所以本着模型简约的原则，也可以将其删除。

第三个问题可以这样理解，C 与 F 的关系是伪相关，被后面加入的变量替代了。举个例子，一个人的身高和他的学习成绩之间成显著正相关，但加入年龄这自变量后，身高这个变量与成绩间的关系变得不显著了。这个和你的这个例子可以类比。

问题四的理解同问题 3。不过同时请注意构建模型时有没有对数据进行预处理，小心极端异常值或者严重多重共线性对结果造成的影响。祝好运~



提问：

HLM 模型问题！

在建立 HLM 文档的时候，在 make HDM 的时候总是提示“missing data found at level 1,unable to continue!"，该怎么解决？难道有缺失数据就不能运行 HLM 吗？我在上面 missing data?已经勾选了 Yes，而且也在 delete missing data when 勾选了 running analyses.

解 答 栏

建议在用 HLM 软件跑模型前，先人为的将缺失值删除或者填充，然后再建模。祝好运~



提问：

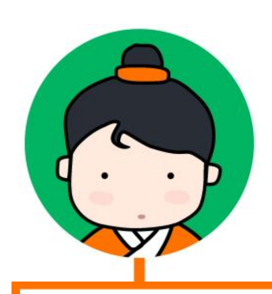
多层次模型中层 2 的数据可以是分年的吗？

准备用多层次模型处理数据。数据是按省来分组的，层 2 的数据有人均 GDP，那么这个人均 GDP 的取值可以包括 2012 至 2014 年的人均 GDP，还是只能是某一年的人均 GDP 数据？

解 答 栏

可以这样理解这个问题，如果你使用的是一套横截面数据（如 2012 年对应的各省的数据），那么你第二层用 2012 年对应的人均 GDP 就行。但我看你的描述，你所使用的数据应该是一套面板数据，那么你可能还需要对时间序列再分一层。即整体模型为三层。具体的原理和介绍推荐看郭志刚等翻译的《分层线性模型：应用与数据分析方法》一书，有相关介绍。链接如下

http://bbs.pinggu.org/thread-1188193-1-1.html。祝好运~



提问：

灰色关联度分析

灰色关联度分析和计量回归模型都能够分析出两个因素之间的相关关系，那么这两种方法在运用上究竟有何不同呢？求高人指点！

解 答 栏

两者之间存在着很大的不同的，http://baike.haosou.com/doc/6090168-6303274.html，这是灰色关联度的介绍，楼主看看后就明白了。祝好运~

6.3.6 期刊投稿

核心期刊论文对格式要求往往比较严格，对于常常只注重论文内容不注意形式的作者们来说，核心期刊论文的格式要求直接影响编辑的审稿印象和成功通过与否，显得格外的重要。

核心期刊论文的格式要求会根据不同的期刊会有所不同，但是绝大部分都是一样的，所谓万变不离其宗，只要掌握了论文发表的基本格式，就算期刊编辑有再复杂严格的格式要求，也能轻松搞定，让论文投递更加有把握。

1、核心期刊论文的标准格式为：

文章标题

作者姓名

作者单位：（包括单位全称、邮政编码）

[摘　要]（以摘录或缩编方式复述文章的主要内容）50～300 字

[关键词]（选用可表达文章主要内容的词或词组）3～8 个关键词正　文

参考文献：[1] [2] [3]……

（一般期刊还要求英文摘要和英文关键词）

作者简介与作者联系方式

2、针对以上格式组成还须注意的是：

一、标题

核心期刊论文题目是一篇论文给出的涉及论文范围与水平的第一个重要信息，也是必须考虑到有助于选定关键词不达意和编制题录、索引等二次文献可以提供检索的特定实用信息。

论文题目十分重要，必须用心斟酌选定。有人描述其重要性，用了下面的一句话："论文题目是文章的一半"。

（1）. 准确得体要求论文题目能准确表达论文内容，恰当反映所研究的范围和深度。

（2）. 简短精炼　力求题目的字数要少，用词需要精选。至于多少字算是合乎要求，并无统一的"硬性"规定，一般希望一篇论文题目不要超出 20 个字.参加医学论文发表相关栏目文章。

（3）. 外延和内涵要恰如其分"外延"和"内涵"属于形式逻辑中的概念。所谓外延，是指一个概念所反映的每一个对象；而所谓内涵，则是指对每一个概念对象特有属性的反映。例如：如何快速发表论文。

二、正文

核心期刊论文格式要求正文篇幅一般在 5000--10000 字不等，包括简短引言、论述分析、结果和结论等内容。文字太少就不能充分展开论述。文中出现的外文缩写除公知公用的首次出现一律应标有中文翻译或外文全称。

文中图、表应有自明性，且随文出现，并要有相应的英文名。文中图的数量

一般不超过 6 幅。图中文字、符号、坐标中的标值和标值线必须写清，所有出现的数值都应标有明确的量与单位。文中表格一律采用"三线表"。

文中有关量与单位必须符合国家标准和国际标准。用单个斜体外文字母表示（国家标准中专门规定的有关特征值除外；如要表示量的状态、序位、条件等，可对该单个字母加上下角标、阿拉伯数字以及"′""＾"等），避免用中文表示。

正文章节编号采用三级标题顶格排序。一级标题形如 1，2，3，…排序；二级标题形如 1.1，1.2，1.3，…排序；三级标题形如 1.1.1，1.1.2，1.1.3，…排序；引言不排序。

三、参考文献

核心期刊论文格式要求有专著（M），论文集（C），报纸文集（N），期刊文章（J），学位论文（D），报告（R），标准（S），专利（P），其他未说明文章（Z）

参考文献如为专著，项目包括：作者姓名. 书名. 版本. 出版地：出版者，出版年；参考文献如为期刊，项目包括：作者姓名. 版本. 年. 月. 卷(期)～年. 月. 卷(期). 出版地：出版者，出版年；参考文献如为电子文献，项目包括：作者姓名. 电子文献题名. 文献出处或网址，发表或更新日期.

四、作者信息

包括作者简介（100 字以内） 出生年月 性别 毕业院校 学历 主要研究方向。作者联系方式，包括： 地址， 邮编，电话，(含手机）E-mail 等。



提 问：

查找文献时，找不到这种数据的统计方法！

各位大神，本人有一批人群（300 人）前后三年的自身对照数据，想分析血清某元素的变化与这批人群 3 年后发生糖尿病（30）的关系，请问用什么统计方法！主要想突出此元素的变化与疾病发生的关系！

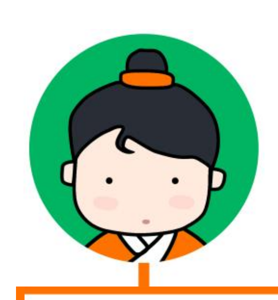
logistic 回归分析可以根据差值的变化，比如分成(group 1：<-2, group2: -2~2, group3: >2）这样三组，探讨与糖尿病的关系，做 logistic 回归吗？高手在哪里？？？

我看许多文献里 Logistic 回归分类后最后一列都有 P for trend 这项检验，请问在 SPSS 上具体怎么操作呢？

解 答 栏

从问题描述来看——“同一批人群前后三年的自身对照数据”，可判定楼主

所用数据为面板数据。因变量疾病发生与否是个类别变量，那么对应的模型方法选择应该用二分类 logistic 回归模型。综合来看，应该使用面板二值选择模型。SPSS 软件可能实现不了，推荐使用 STATA 软件，使用 xtlogit 命令。推荐陈强老师《高级计量经济学及 STATA 应用》一书，上面有该方法的原理介绍，代码实现介绍和案例介绍。祝好运~



提 问：

评审人意见如下,看不懂呢！

一个文章，接收到了审稿意见，评审人提出了如下问题。

我连问题都看不懂，特此求助啊。

谢谢各位同行！真心感谢！

(2) Your data have clusters and thus IID violations. This requires the use of cluster-corrected standard errors or possibly even HLM. Since most regression packages allow for cluster-corrected SEs only on one

level, please show the ICC(1) values for each possible area of clustering (in particular, industry, region).

(4) As a robustness test, please include in the paper results using a Tobit regression specification. Again, this will require cluster-corrected SEs.

解 答 栏

楼主应该是用的面板数据吧，在做回归时用的是一般的标准误。审稿人的意思是要给出 cluster(id)的稳健标准误。同时，楼主的变量应该是分层的变量（如农户个人，家庭，村落这种是有层次的变量），为了消除不同层间的异质性，审稿人建议楼主用 HLM 软件做层次线性模型分析。最后一个问题审稿人是让楼主做 Tobit 回归，但回归时要用稳健标准误 cluster(id)。看来楼主是遇到懂行的审稿人了，祝好运。

同意 xddlovejiao1314 对第二个修改意见的说明。

对第一个修改意见，没建议一定要用 HLM 软件做，只说“or possibly even HLM”，表明 HLM 可用可不用。

第一个修改意见，主要在 please show the ICC(1) values for each

possible area of clustering (in particular, industry, region). 不懂 ICC(1)的意思，不知是不是分别显示 vce(cluster industry)、vce(cluster region)等估计结果。

这个审稿意见不仅指出了问题，还针对问题给出了具体的修改意见，很难碰到，楼主真幸运。

追 问：

好人啊~请问这个怎么做呢？通过 SPSS 能实现吗？唉，没弄过啊。我的就是截面的问卷数据，但的确可能涉及到跨层次的问题。好像是这个意思，如果不用 HLM，那用什么实现审稿人说的呢？

除了 HLM 软件之外，STATA 也可以做分层线性模型。推荐汉密尔顿著，郭志刚等翻译的《应用 STATA 做统计分析》一书第 15 章“多层与混合效应建模”，该章节专门介绍了这种方法。此书的电子版链接如下：http://bbs.pinggu.org/thread-944900-1-1.html。同时，还可以查查 STATA 的帮助手册，里面应该也有介绍。此外，如果要用 HLM 软件实现的话，推荐

张雷老师的《多层线性模型应用》一书（http://bbs.pinggu.org/thread-3035963-1-1.html），该书不仅有多层线性模型的简单原理介绍，还有 HLM 软件操作截图，在此书的指导下能快速的得到想要的结果。不过，如果要掌握好多层线性模型的原理，推荐郭志刚老师翻译的

《分层线性模型：应用于数据分析方法》一书

（http://bbs.pinggu.org/thread-1188193-1-1.html）。祝好运~



提 问：

发表论文，是要先选择目标期刊呢，还是先把论文写好再找目标期刊呢

发表论文，是要先选择目标期刊呢，还是先把论文写好再找目标期刊呢

解 答 栏

人与人之间存在差异，所以我个人觉得这个问题不能一概而论，应该视具体情况而定。就我个人而言，我前期是先写好文章，后找目标期刊。即有 idea 先把文章写好，然后到知网、谷歌学术等地方以文章关键词进行检索，去找目标期刊。这样做可能有一点不好，就是早期看的文献可能期刊质量一般，所以写出来

的文章质量也一般，拿在手里比较尴尬；现在写文章基本会先锁定几个比较好的目标期刊，然后有针对性的看某方面的文献，最后再写作投稿的。其实，现在看来，不管是哪种方法都各有其优劣，应该视个体具体情况而定，可能两者都考虑一下为佳。祝好运~



问在经济学的路上

七、

问在经济学的路上：真正的学习从提问开始

〖从事设计专业的统计小白被迫走上学术道路，然后面对一堆建模数据和方法选择无从下手，就连最基本的数据描述性分析都不会。误打误撞中来到该论坛的 SPSS 专版提下了一个现在看来很白痴的问题，本来以为会跟其他地方一样没人理睬，但是天降 xddlovejiao 大神热情地对我的问题详细的进行了解释，还推荐了很多专业书籍，让我一步步对 SPSS 和逻辑回归分析有了初步的了解。而且最让人敬佩的是，大神在帖子中不嫌麻烦地针对逻辑回归的原理/数据处理方法等内容，一步步地跟我解释，回答速度也非常快，大有版聊之势〗

经管之家答疑解惑感言集

由于毕业论文里的计量模型不会做，时隔一年我不得不再次登录人大经济论坛来寻找这方面的专业人士帮我解答疑惑。我在人大经济论坛里的计量经济学与统计论坛中 SPSS 专版里寻找着我想问咨询的问题，我认真的看着每一个回复者的精彩解答，我发现一个 ID 名叫 xddlovejiao1314 的学长，对于每个问题，都会很耐心的回答，并且回答的很详细。于是我就试着向他提问，我把我的困惑发给他，问的问题很基础，只是发给了他我论文的变量数据，请教他用什么模型比较好，我以为这么简单的问题会被他忽略，没想到他很快就回复我了，并且回复的很详细，让我很兴奋，于是我再继续追问他：那么我采用岭回归可以吗?他给出了的建议是：不太建议用这个方法消除多重共线性，可以用主成分分析。并且主动温馨提醒我，我变量的选取有问题，应该这么选择。我跟徐师兄继续回复，我电脑没有 SPSS 软件怎么办呢？于是他主动加我 QQ 号，把 700M 的 SPSS 安装软件发给我，传输花了 2 个多小时。在一个陌生的环境中，徐师兄非常有耐心的回复我每一个问题，让我很感动。

接下来的几天，我感觉他就是一位我论文模型的救星，我计量经济学基础很差，很多不懂，做出的模型都会发给他看，他也很快就回复我，帮我指出模型的意义，所以我的毕业论文模型算是在他的帮助下一点点做起来的。但是徐师兄从来不会主动替我做模型，给我的是建模思路，同时还会把那些相关资料一并发给我，让我自己亲自去完成。感谢他在百忙之中抽出宝贵的时间无偿为我解答太多疑惑，非常感谢！

希望徐师兄在未来的学术道路上越走越远！

晨光熹微 2013

2015.8.5

只会 SPSS 基础操作的小硕以全国性数据写毕业论文真心不易。从没有发过求助贴的我第一次在人大经济论坛求助。当小博士热心的回复我之后，我还是有些不能理解，所以也是第一次问陌生人的联系方式，希望可以多学习，本以为对方会觉得很荒谬而不再帮助我，没想到小博士告诉我了 QQ 号和手机号码，让我瞬间觉得人与人之间并不是都那么冷漠。

小博士在网上一步步的教我如何操作，包括对数据的筛查，以及统计指标的解释让我对数据分析多了一些想法 然后自己会明白从哪里入手分析繁杂的数据，当我看到 SPSS 输出结果的时候，也会觉得数据是很神奇的存在。当然也是在小博士的指点下，意识到自己 SPSS 破解版软件不稳定，解决了 SPSS 数据分析不一致的困惑。

最后希望小博士在学术的道路上越走越远。

anran19921207

2015.8.5

从事设计专业的统计小白被迫走上学术道路，然后面对一堆建模数据和方法选择无从下手，就连最基本的数据描述性分析都不会～～～误打误撞中来到该论坛的 SPSS 专版提下了一个现在看来很白痴的问题，本来以为会跟其他地方一样没人理睬，但是天降 xddlovejiao 大神热情地对我的问题详细的进行了解释，还推荐了很多专业书籍，让我一步步对 SPSS 和逻辑回归分析有了初步的了解～～～而且最让人敬佩的是，大神在帖子中不嫌麻烦地针对逻辑回归的原理/数据处理方法等内容，一步步地跟我解释，回答速度也非常快，大有在帖子中版聊的趋势～～

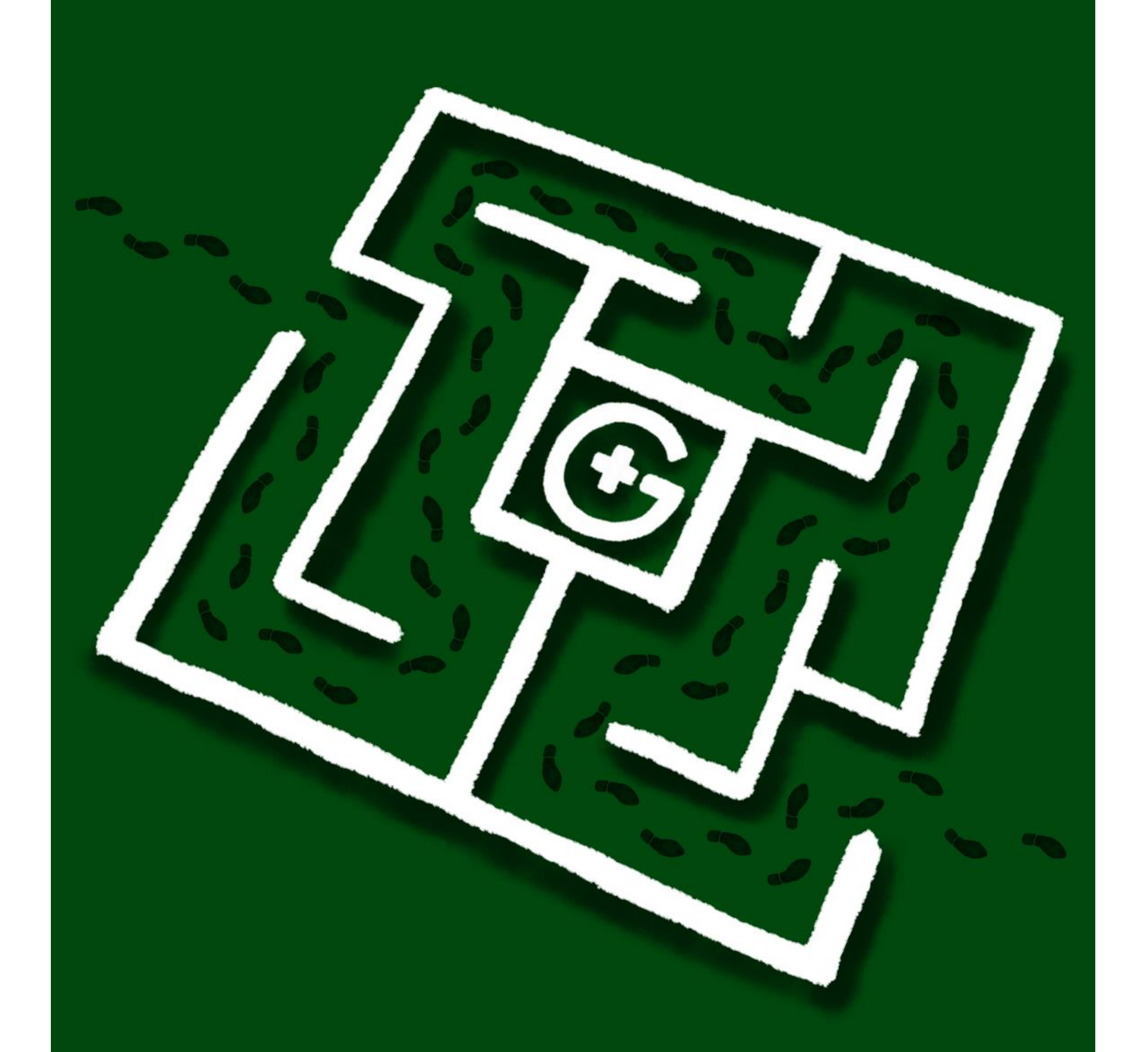
在他的帮助下，完全没有统计知识的我现在已基本对逻辑回归中的操作方法和原理有了一定 的了解，并且还产生向更复杂的 STATA 软件进攻的动力，这样就可以完成大神推荐过的缩尾处理了～～

因为曾经在其他论坛上碰到了问题没人回答的冷遇，所以能够在这里得到大神这么及时有效的回复，真的很让人感动，待我毕业论文大功告成的那天，一定得在致谢中好好表扬下大神的帮忙啊～～～

ajiao4310 坛友

xddlovejiao1314——我要非常感谢他，我在使用 SPSS 时遇到一些自己拿不准的问题，主要是自己不清楚因子分析最终结果是怎么计算出来的，如何求指标之间的相关系数阵？哪个表是成分得分系数矩阵？因子分析模型中系数应该使用哪个表个数据？xddlovejiao1314 一步一步地给我解答，解答中途遇到新问题（这主要是自己无知），xddlovejiao1314 都耐心帮助我，而且解答非常及时。非常感谢 xddlovejiao1314，祝工作顺利。

一位人大的学生


为每个经管人解决学习难题

八、

为每个经管人解决学习难题：在你身边 为你答疑

〖知识的价值不在于所有权，而在于它能够让更多人更好地理解这个世界，让人们用更美好的方式来生活。聚焦每个经管人的知识，为每个经管人所用，答疑是学习，也是交流。很多时候，答题与其说发现世界，不如说发现自我。答题是一种经历、知识、智慧的梳理与总结，是个人知识积累的原创公开作品集。我们发起并组织经管答疑，为你，为我，为解决难题，为知识分享，也为智识传递……〗

经管之家（原人大经济论坛）之经管答疑

学贵知疑， 知识的价值不在于所有权，而在于它能够让更多人更好地理解这个世界，让人们用更美好的方式来生活。

聚焦每个经管人的知识，为每个经管人所用。答疑是学习，也是交流。很多时候，答题与其说发现世界，不如说发现自我。答题是一种经历、知识、智识的梳理与总结，是个人知识积累的原创公开作品集。我们发起并组织经管答疑，为你，为我，为解决难题，为知识分享，也为智识传递……

——经管答疑：为经管难题而来 为解决难题而生

亲爱的看官，您面前这本小小的电子书，看到这儿已近尾声了，但你知道吗？它的缘起，其实是经管之家的经管答疑。我们没有想到，有一部分的答疑成果，会以书的形式与您见面；我们更没有想到，江湖之高，庙堂之远，会有那么多热心的经管人参与经管答疑。回答，分享，反馈，碰撞，答案优化，就这样，一个

个知识锦囊热心传递下去，一个个零回复难题得到有效解答……

经管之家（原人大经济论坛）走过了十年，有将近 700 万的注册用户，积累了将近 3000 万个贴子。2015 年起，我们开始组织经管答疑，邀请经济、管理、金融、计量统计等各领域的专家、版主、热心坛友投入答疑。每天，都有各路专家穿梭在网站各个版块，发现问题，回答问题，整理答案，听取提问者的反馈。从最初的不足十人，到今天，已有对经管知识有强烈兴趣和分享意识的多位经管人，特别是经管之家的近 20 位专业版主，共同投入了这场答疑大战，这场知识共济和分享互助的旅程。

9 个月来，经过我们的初步整理，各学科的已回答的精彩问答主题贴就将近 2 万个，历年来精彩问答超过 2 万个。这意味着，在经管学习的旅途中，您踩过的坑，跳过的陷阱，碰到的难题，别人也一样遇到过。如今，您可以沿着前人的提问和回答的脚步，更快捷地找到答案，更好地与经管难题触手而及……

在参与经管答疑的过程中，大家渐渐发现，对自己专业领域问题的回答其实是一个很好的巩固学习的过程。因为只要你认真地去回答一个问题，你必然会力求自己的表述是客观的、正确的、有逻辑的、有条理的。这是帮助你梳理知识、归纳知识、沉淀知识的好方法。于是，各路英雄积极参与，又反推了提问者的热情。我们粗略估算，论坛中积存的经管问答，超过十万个，一个全新的经管领域

的十万个问为什么，正在呼之欲出。

十年来，互联网领域的变化天翻地覆。如今的信息，不是过少了，而是过剩了。经管领域聚焦了很多内容网站，却没有一个专注、专业的经管知识答疑平台。于是经常有人感慨：论坛的内容如此丰富，简直是经管人学习和工作中的宝藏；但论坛的内容太过丰富，让提问的问题倏忽间就被淹没、让找答案的人费尽心力最终找出的内容却差强人意。

今天，经管之家之“经管爱问”系统，就是我们为解决这一问题付出的努力。大道至简，知识之花常开！我们希望从论坛丰富的积淀中提炼、塑造出一个专注而专业的答疑平台，为那些想知道答案和想分享知识的经管人而生。

如果你也有经管难题，请来经管之家提问。

如果你正是为了传道授业解惑而来，我们需要的，正是你这样的人......

——不答不相识：答疑团队与答疑成果

为了让小伙伴们得到更好的答案，我们一直在竭尽所能。

经管世界，财经江湖，高手辈出，甚至有很多是深藏的大侠，需要把他们挖掘出来。于是，我们邀请了这样一群答疑专家，典型的他们是这样的：中国和北美双精算师、统计软件高手、课题研究者、清华状元、美国商学院教授、纽约联

储局研究助理、国内前十大会计师事务所专员、国内一线教授和学者、编程高手、国内外优秀经管硕士和博士……他们不仅有这些熠熠生辉的标签，还是在各个领域扎扎实实、勤勤恳恳的经管人，他们严谨、专业而友好，期待和大家有更多的沟通和互助。最重要的，他们努力、尽心、细致地答疑，一心帮助大家找到最好的答案。

经管答疑第一小分队，是我们聘请的官方答疑编辑老师，包括 crystal1882、matlab-007、学院派福尔摩斯、胖胖小龟宝、bakoll、见路不走、午夜的太阳和 yangyuzhou 共 9 位老师，他们每天定时巡视自己负责的方向，确保问题第一时间得到关注和解决；第二梯队是答疑版主及特邀专家，包括 whe58、lwzxy、dreamtree、wwqqer、长风神舞、xddlovejiao1314、福荣山、fangzhao、accumulation、oliyiyi、李攀、夏目贵志、铁锷未残、STATAx、jiangbeilu 和不死稻草人等经管之家的优秀版主，他们同样会定期巡视自己负责的学科版面，对版块所在的提问进行补充回答和扩展回答；同时，我们努力通过组织发起、灵活参与等方式，让经管世界中具有知识盈余但时间不多的坛友以自己的方式竭尽全力参与到答疑中来；最后，我们最强大的内援团队，就是全体坛友，虽然以上的答疑团队覆盖面已经很广，但我们相信如果没有群众的力量，看起来再坚固的“马奇诺防线”也可能被轻松绕过。我们会通过答疑财富值、答疑奖金池计划等活动，让更多的坛友参与进来，不管是在经管爱问系统、爱问频道还是其它专业

版面，如果您的答疑被我们的客服人员巡视到，都会得到超额的论坛币奖励。也许这对您来说是举手之劳，但我们因有了大家的点滴分享和帮助而更完整。

我们组织的的多维答疑团队，能够解决多类经管和财经领域的难题。目前，常见的计量与统计理论及软件、实证文章写作、金融学、投资学、风险管理、量化投资、微观/宏观/产业/区域/制度/能源/环境等经济学领域、精算学、财务与会计、人力资源/市场营销/战略管理/管理学基础/组织管理等工商管理领域、国际贸易与国际商务、行业研究以及数据清洗等问题多发的领域。我们已经在这些领域回答了 20000 多个问题，同时我们的团队成员还整理了将近 20000 个论坛中已经回答的问题，这些数字每天都在增加，我们希望竭尽全力解决论坛中的每一个经管问题。如果您想详细了解答疑老师及其答疑范围，请访问答疑专家信息列表：http://ask.pinggu.org/?expert/default.html，您可以随时把您的问题链接发给他们，以便得到更好的解决。

我们是一个怀着理想主义信念的团队，希望让知识的血液在互联网的空间中无限的畅流、永续地传承，让每一个经管和财经问题从此都有解。我们相信，分享不是慷慨，分享是明智的，它终究会让我们得到更多。每天坚持答疑，有时候的确很枯燥，难以想象怎样的自信与坚持才能够撑起那些不为人知的平凡与琐碎？在此向每一个参与答疑的团队成员表达我们最真诚的感谢与敬意。

——所有的提问都是为了更好的学习："经管问吧"与"经管学吧"

教学相长，学习的本质从提问开始，提问本身最终会为学习服务。

经管答疑之"经管爱问"系统首先是一个"问吧"，一个经管、财经领域的专业、社交、互动问答平台。你的问题总有人知道答案，我们通过这一互联网平台对接问答双方，为知识的需求者和提供者搭建信息获取的桥梁，从而构建成一个更为集中、专业、垂直的知识分享生态链，相信通过众人的帮助、分享和互动，您的问题终能得到一个合适而优质的答案。

学以致用，用则加深。我们相信问答如果不能内嵌到一个系统中，就会在丰富繁多的信息流中淹没；知识如果不能深刻的内化为我们的思想，就会很快离开我们。所以我们将优秀的知识问答梳理、归类、提炼，甚至总结成专题和电子书，让它成为一个"经管学吧"。目前，我们的学吧已经分为经济、管理、金融、计量与统计理论、计量软件、数据分析与挖掘、考研考博、考证等 16 个在线答疑学习模块，并不断扩容。通过自动化和人工的方式分拣，将不同的知识归类整理以便成为一个体系，方便用户更好地化知识为思想、化思想为行动的力量。

这是一个为经管人而生、为经管问题而生的平台，你可以是经济、管理、金融、统计、财会等各专业的学生，也可以是经管职场人，甚至你可以只是对其中某一领域充满好奇的经管边缘人，这里都是为你而设计，你可以问、可以答，也

许你很久才提问一次，也许你只是偶尔回答了一两个问题，我们都因为有你而更加完整。这是一个新生的平台，我们愿意和每一个经管人一道，一起呵护和扶持它，让它成长为一个更优质的经管答疑与知识管理平台。

——你的经管难题怎样丢过来？：经管之家经管爱问、问答微信、人大经济论坛 APP、问答微博、助答 QQ 群无所不在

无论你在哪里提问，无论你在电脑还是手机上。只要你在，我们就会解决你的经管难题。

经管之家经管爱问答疑专区：在论坛“经管爱问”系统发贴提问，我们的答疑老师会每天巡视，竭尽全力解答每一个问题。如果您仍然担心问题会被忽视，可以将问题链接用站内消息发送给账号“论坛答疑”，我们的客服人员会优先处理。

答疑微信：您可以关注公众微信平台“经管爱问”（jgasker)，通过每日推送、底部工具栏、直接留言等方式获得更快捷方便的移动答疑和学习体验。

人大经济论坛 APP 答疑：在各大应用市场下载“人大经济论坛”app，其中内置了经管答疑模块，可以直接用手机阅览和发贴提问，随时随地和论坛内容同步。

微博和 QQ 群答疑：微博用户在新浪微博搜索“人大经济论坛_经管答疑”

即可找到我们，千问千寻，和博主一起探索经管答疑世界；另外，欢迎大家加入人大经济论坛官方助答 QQ 群（237164035），用更自由的方式和众多经管人一起探讨互助。

BTW,我们还想再爆个料，其实咱家还有两位专属的答疑客服人员，每天随时巡视各入口的经管难题，随时反馈专家，或直接为您回答哦。是不是很快呢？当然，有问有答！第一时间努力回答！

经管答疑，你手里有熊掌，我手里有鱼，分享一下，兼得.......

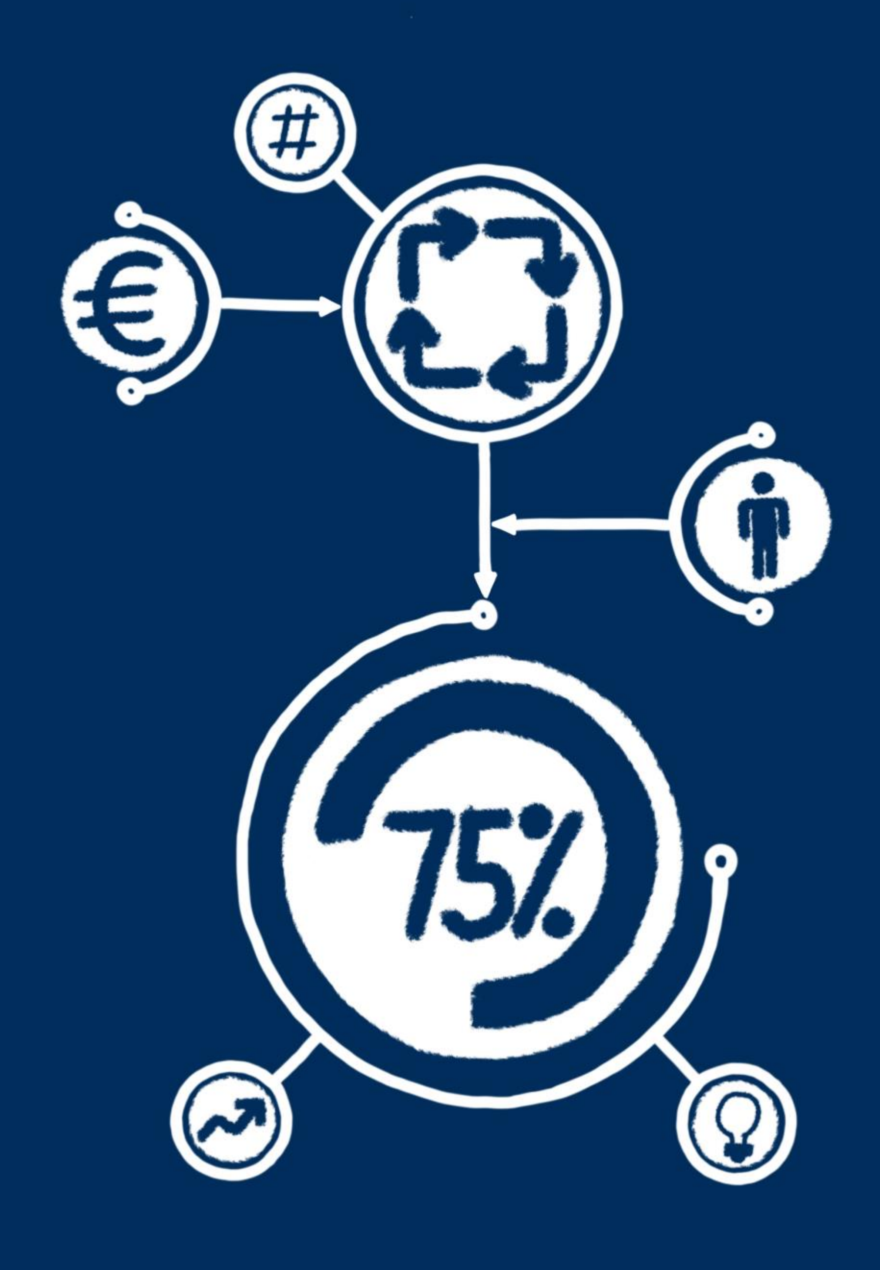
经管答疑，聚焦每个经管人的知识，为每个经管人所用......

有经管难题来经管之家，每天在这里找到 N 个问题的答案。

经管爱问，为经管人和经管问题而生！为更有品质的经管学习而来！

经管之家经管答疑团队

2015 年 12 月



与大数据共眠

九、

与大数据共眠：经管之家大数据与数据分析好课不断

〖如今，我们已经进入了企业发展日新月异的“互联网+”时代----一个用数据说话的时代，也是一个依靠数据竞争的时代。数据的分析和处理能力正在日益成为企业倚重的技术手段，对数据分析方面人才的需求更加迫切〗

9.1

CDA 数据分析师：学习是最有效的投资

CDA 数据分析师走过的这两年



2006 年，人大经济论坛数据分析培训中心设立，至今经历 9 个春秋，建立了大陆、台湾一线师资团队，培养人才已达 3 万余人；

2013 年，“中国数据挖掘与数据分析俱乐部 CDMC”在人大经济论坛旗下成立，2014 年改名为“中国数据分析师俱乐部 CDA”。来自政府、金融、电信、零售、电商、互联网、教育等行业人士加入会员，成功举办了近二十场行业聚会；

2013 年，CDA 数据分析研究院成立，CDA 数据分析师课程第一次走进大家的视野，截至到现在，累计培养 1200 多名专业数据分析师；

2015 年，CDA 就业学院成立，响应国家职业教育人才培养的号召，计划与企业对接，在全国高校培养数据分析专业人才。截至到目前，我们帮助在 200 多个高校成立了大数据俱乐部和兴趣小组。

2016 年，精彩在继续……

CDA 数据分析师这个品牌自成立起，就致力于帮助企业人员建立系统化的分析方法和思路，提升他们在商业数据分析领域的能力和洞察力。

（CDA 数据分析师是指在互联网、零售、金融、电信、医学、旅游等行业专门从事数据的采集、清洗、处理、分析并能制作业务报告、提供决策的新型数据分析人才。CDA 数据分析师覆盖了国内企业招聘数据分析师所要求的所有技能，包括统计知识、软件应用（SPSS/SAS/R 等）、数据挖掘、数据库、报告撰写、项目经验等。CDA 数据分析师分为 LEVEL ⅠⅡⅢ三个等级，成为一名合格的 CDA 数据分析师能够胜任企业不同层次的数据分析工作。）

这两年，我们整个团队在做的东西，总结起来有三件事情：

一是总结论坛过去上百门课程的经验，提取课程精华，完善整个培训体系，把大数据这个相对比较高的概念降低下来，让我们的学员踏实去学习一些基础性、系统性的理论和方法。

具体来讲，当时我们从各大招聘网站上面抓取了众多数据分析师岗位的需求信息，去了解企业需要什么样的数据人才，从很多中小企业的情况来看，一是数

据分散，尚没有分析体系，二是分析人员岗位体系还未成熟。从 CDA 数据分析研究院出来的分析师，能够结合具体的业务问题，运用系统有效的分析方法，帮助企业去完善自己的分析体系。



主流数据分析软件如 SPSS、SAS、R 等的培训，是我们过去 9 年来最擅长的领域，结合积累的大量培训经验和丰富的教材案例，我们将软件应用能力作为课程的核心要求之一，贯穿整个授课周期，让每个人都精通专业的数据分析工具。在此基础上，后续课程对商业理解、业务分析思路的培养，则是将每一个人的能力体系真正建立了起来。我们认为，数据分析工作终究是服务于实际业务，一盘菜上来了，大家关注的是好不好吃，至于是用微波炉还是铁锅做的，厨师有自己的选择。CDA 数据分析师的宗旨就在于助力商业数据分析人才。

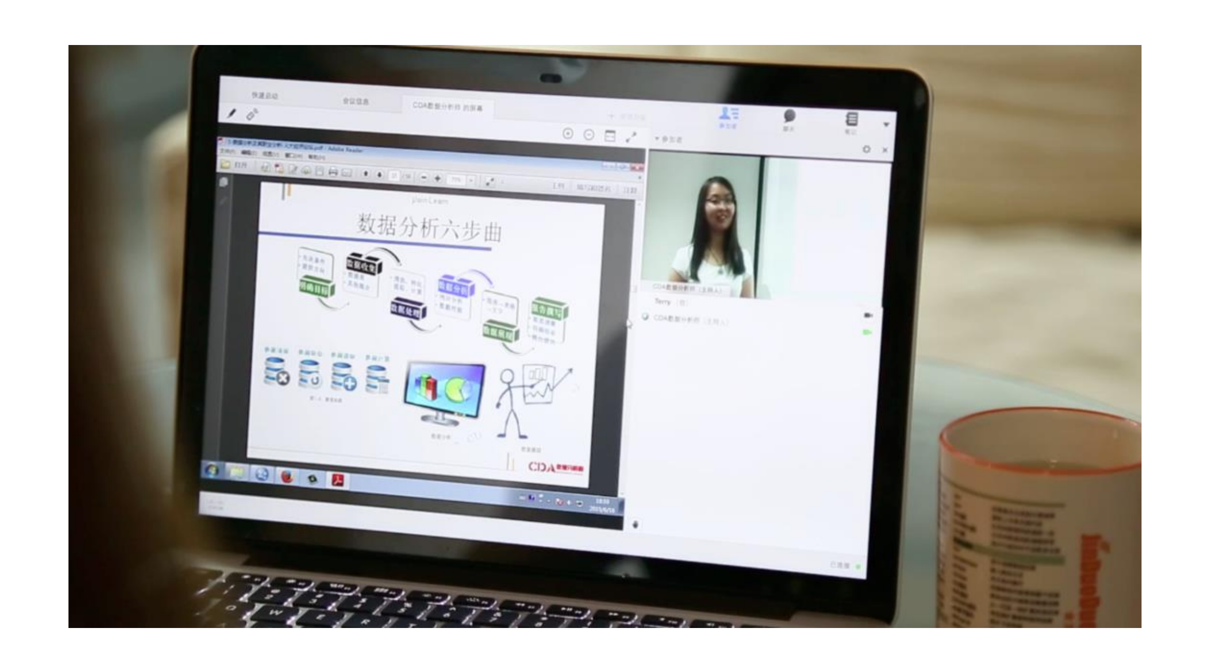
目前 CDA 数据分析师已经推出了 Level1 和 Level2 两个等级，三个方向的课程，以真实的工作环境为依据，融合理论基础，将真本事融会贯通教授给学员，

在 CDA 大数据课程和 CDALevel2 课程中，我们提供了一整套的搭建数据分析平台的思路和具体方案，这是在互联网公司极为领先的解决方案。与此同时，我们也一直在探讨 Level3 即数据分析专家的培养考核。

各个等级的能力要求和定位如下：

等级要求	Level Ⅰ 业务数据分析师	Level Ⅱ 方向一：建模分析师	Level Ⅱ 方向二：大数据分析师	Level Ⅲ 数据分析专家
行业覆盖	◆ 专指政府、金融、电信、零售等行业前端业务人员 ◆ 从事市场、管理、财务、供应、咨询等职位的业务人员	◆ 专指政府、金融、电信、零售、互联网、电商、医学等行业专门从事数据分析与数据挖掘的人员	◆ 专指政府、金融、电信、零售、互联网、电商、医学等行业专门从事数据分析与数据挖掘的人员	◆ 专指从事各行业、企业整体数据资产的整合、管理的专业人员，面向用户数据创造不同的产品与决策，一般指首席分析师（CA）
工作经验	◆ 非统计、计算机专业背景零基础入行和转行就业人员	◆ 两年以上数据分析岗位工作经验 ◆ 或通过CDA Level Ⅰ认证半年以上	◆ 两年以上数据分析岗位工作经验 ◆ 或通过CDA Level Ⅰ认证半年以上	◆ 五年以上数据分析岗位工作经验 ◆ 或通过二级认证半年以上
理论基础	◆ 掌握概率论和统计理论基础	◆ 在Level Ⅰ的基础上更要求掌握多元统计、时间序列、数据挖掘等理论知识 ◆ 掌握高级数据分析方法与数据挖掘算法	◆ 在Level Ⅰ的基础上要求掌握JAVA语言和linux操作系统知识	◆ 掌握CDA Level Ⅱ的所有理论及技术要求，大数据分析架构及企业战略分析方法
软件技能	◆ 熟练运用Excel、SPSS、SAS等一门专业分析软件	◆ 熟练运用SPSS、SAS、Matlab、R等至少一门专业分析软件 ◆ 熟悉使用SQL访问企业的数据库	◆ 掌握运用Hadoop、Spark、Storm等至少一门专业大数据分析软件	◆ 了解计算机技术，软件开发技术
能力贡献	◆ 良好的商业理解能力，能够根据业务问题指标利用常用数据分析方法进行数据的处理与分析，并得出逻辑清晰的业务报告	◆ 结合业务，能从海量数据提取相关信息，从不同维度进行建模分析，形成逻辑严密能够体现整体数据挖掘流程化的数据分析报告。	◆ 从海量数据中提取相关信息，并能够结合R python等软件，形成严谨有效的数据分析报告	◆ 带领团队完成不同主题数据的有效整合与管理。对行业、业务、技术有敏锐的洞察力和判断力，为企业发展提供全方面的数据支持

在与众多企业的接触过程中，我们也深刻感受到企业对于这几类人才的强烈需求，不管是业务部门，还是 IT 部门，都是希望有分析素养的人才加入进去的。目前，我们同近百家企业的 HR 保持充分的沟通，互相推送岗位需求和优秀学员信息。



第二件事情，是完善我们的课程体验。包括现场课程服务以及在线直播技术的运用。在 CDA 课程的第三期，我们有这样一位来自于武汉的学员，每周都来回于北京和武汉之间，利用工作间隙，完成了跨时 8 天近 60 个课时的学习；我也接触过陕西某部队的基层军官，一位数据分析的爱好者，千里迢迢过来参加我们的技术聚会......即便是在数据分析岗位、分析文化并不是很完善的二、三线城市，也同样汇聚了一大批的数据分析爱好者。2014 年初，我们跟思科合作，利用直播平台 Webex 的领先技术，形成现场课程的高清实况影像与清晰的教学演示，把 CDA 课程的优质内容以及汇集的优秀师资资源，推送到更多地区，越来越多的会员在加入我们的在线学习课堂，与现场学员一起，与我们的老师、助教一起，学习和探讨问题。

学习向来不是件容易的事情，在优化课程内容的同时，我们又在现场环境和服务方面作了一些探索，我们的工作人员在周末经常会亲手制作一些咖啡茶点，布置好舒适的学习、休息环境。愚人节那天，我们跟会员们开了这么一个玩笑：应聘的人还挺多，其实我们是想告诉大家，我们的工作人员们是很温柔，很细心负责的。

我们一直在做的第三件事情，在我们看来，是公益与商业的结合，是能够影响大数据行业生态，具有深远意义的。我们一直在建立和完善学习社区，营造一个数据分析爱好者的家园。

论坛集百家之言，汇四方资源，一直都在创造和传播优质的经管教育资源。各类学习资料，学习经验，学科探讨，都为数据分析爱好者提供了支持。目前论坛各数据分析相关的版块，每天累计有上万人浏览以及讨论相关专业问题，我们目前正在完善经管答疑的学习系统，其中很重要的一块就是 SPSS/SAS/R 语言/MATLAB/EVIEWS/Python/EXCEL 等统计与数据分析挖掘类软件以及商业智能、大数据等相关领域。我们从问答系统中反馈评价非常高的精彩问答中优选内容，加以归类和整理，以答疑学习系统的形式展示出来，同时，也用答疑专题、专刊等表现出来，让大家更易于阅读。

2014 年底，我们又总结往期培训经验，汇总优秀课件，公开出版了 CDA 数据分析师系列书籍 3 本，为菜鸟数据分析师提供理论、技术、经验方面的指导。2015 年下半年，大家期待的后续书籍也会陆续上市。

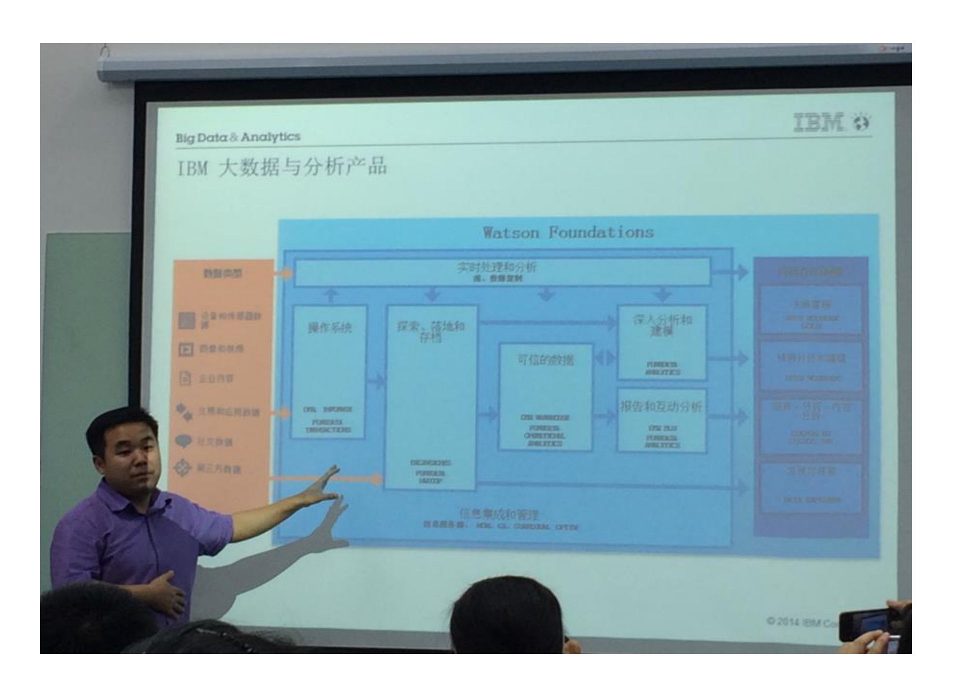


为了丰富众多数据分析爱好者的线下活动，2013 年，“中国数据挖掘与数据分析俱乐部 CDMC”在论坛旗下成立，2014 年改名为“中国数据分析师俱乐部 CDA”。来自政府、金融、电信、零售、电商、互联网、教育等行业总计 920 名人士加入会员，陆续在北京、上海、广州、深圳、成都、河南等地区建立了分会，成功举办了近二十场场行业聚会。CDA 俱乐部定位于建立技术与需求的对接平台，创意与灵感的展示平台以及数据人的朋友圈，从今年开始，北京的会员会

更是以“CDA 数据分析师驿站”的形式，每个周六的相聚在一起，坐下来喝喝咖啡，谈谈合作，听听数据界专家牛人的真知灼见。







2015 年初，我们又成立了 CDA 就业学院，希望给在校的大学生群体提供一个学习和交流的平台，缩小高校传统理论教学和企业实际业务能力要求之间的差距。截至到目前，我们帮助在 200 多个高校成立了大数据俱乐部和兴趣小组，后续将会致力于国内外优秀书籍翻译、参加各类数据分析、数据挖掘竞赛以及校园主题活动的举办。

2015 年下半年开始，CDA 校园合作计划和企业合作计划将会在更多地方推行开来，我们将依托论坛资源优势，为高校提供完整成熟的 CDA 视频学习服务和学分合作计划，为企业提供项目咨询合作、部门组建，员工技能培训等服务。

CDA 研究院是开放的，也是在不断积极改变的，论坛不仅是论坛，CDA 也不仅是培训，更多精彩，敬请大家期待……

最后，CDA 有一句话，我们希望让每个人都知道：真本事，心舒适。

9.2

CDA 数据分析师项目：join and learn!joy in learn！

http://cda.pinggu.org/view/1.html

企业客户


















































什么是 CDA?

CDA 数据分析师是指在互联网、零售、金融、电信、医学、旅游等行业专门从事数据的采集、清洗、处理、分析并能制作业务报告、提供决策的新型数据分析人才。CDA 数据分析师覆盖了国内企业招聘数据分析师所要求的技能，包括统计知识、软件应用（SPSS/SAS/R 等）、数据挖掘、数据库、报告撰写、项目经验等。CDA 数据分析师分为 LEVEL Ⅰ Ⅱ Ⅲ三个等级，成为一名合格的 CDA 数据分析师能够胜任企业不同层次的数据分析工作。

CDA(Certified Data Analyst)，全称“CDA 数据分析师”，由“Certified Data Analyst Institute”（以下简称“CDA 协会”）联合经管之家（原人大经济论坛）在顺应大数据、云计算的潮流下发起成立的职业简称。旨在加强国内外乃至全球范围内正规化、科学化、专业化的数据分析人才队伍建设，进一步提升数据分析师的职业素养与能力水平，促进数据分析行业的高质量持续快速发展。CDA 数据分析师项目包括教育，咨询，考试，认证，机构招聘合作。CDA 数据分析师分为

三个等级（参考认证标准），经管之家每年举办两次等级考试（参考考试安排），通过考试者可以获得 CDA 数据分析师等级证书，此证书代表数据分析师人才技能水平，为企业事业单位选拔和聘用专业人才的参考依据。

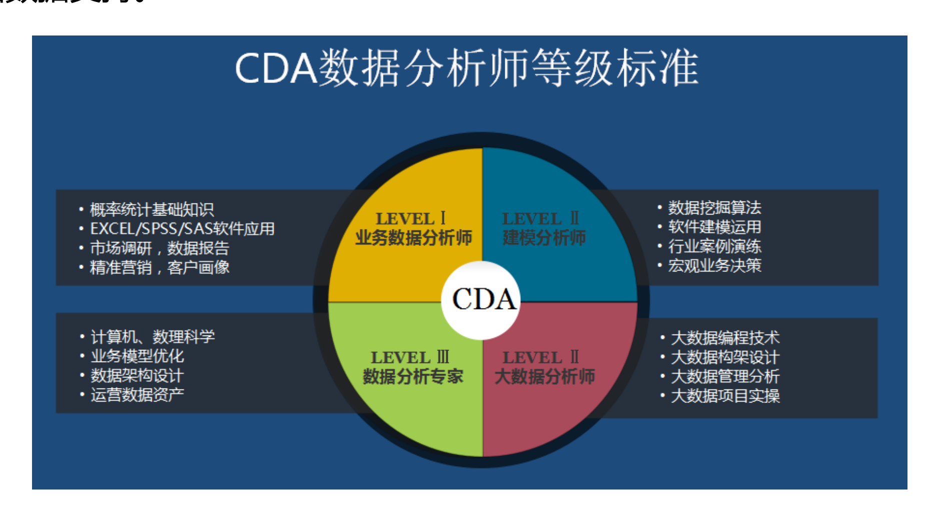
CDA 数据分析师等级有哪些?

CDA Level Ⅰ：业务数据分析师。专指政府、金融、电信、零售等行业前端业务人员；从事市场、管理、财务、供应、咨询等职位业务人员；非统计、计算机专业背景零基础入行和转行就业人员。CDA Level Ⅰ业务数据分析师需要掌握概率论和统计理论基础，能够熟练运用 Excel、SPSS、SAS 等一门专业分析软件，有良好的商业理解能力，能够根据业务问题指标利用常用数据分析方法进行数据的处理与分析，并得出逻辑清晰的业务报告。

CDA Level Ⅱ：建模分析师。两年以上数据分析岗位工作经验，或通过 CDA Level Ⅰ认证半年以上。专指政府、金融、电信、零售、互联网、电商、医学等行业专门从事数据分析与数据挖掘的人员。在 Level Ⅰ的基础上更要求掌握多元统计、时间序列、数据挖掘等理论知识，掌握高级数据分析方法与数据挖掘算法，能够熟练运用 SPSS、SAS、Matlab、R 等至少一门专业分析软件，熟悉使用 SQL 访问企业数据库，结合业务，能从海量数据提取相关信息，从不同维度进行建模分析，形成逻辑严密能够体现整体数据挖掘流程化的数据分析报告。

CDA Level Ⅱ：大数据分析师。两年以上数据分析岗位工作经验，或通过CDA Level Ⅰ认证半年以上。专指政府、金融、电信、零售、互联网、电商、医学等行业专门从事数据分析与云端大数据的人员。在 Level Ⅰ的基础上要求掌握JAVA 语言和 linux 操作系统知识，能够掌握运用 Hadoop、Spark、Storm 等至少一门专业大数据分析软件，从海量数据中提取相关信息，并能够结合 R python 等软件，形成严密的数据分析报告。

CDA Level Ⅲ：数据分析专家。五年以上数据分析岗位工作经验，或通过二级认证半年以上。专指从事各行业、企业整体数据资产的整合、管理的专业人员，面向用户数据创造不同的产品与决策，一般指首席分析师（CA）。数据分析专家需要掌握 CDA Level Ⅱ的所有理论及技术要求，还应了解计算机技术，软件开发技术，大数据分析架构及企业战略分析方法，能带领团队完成不同主题数据的有效整合与管理。对行业、业务、技术有敏锐的洞察力和判断力，为企业发展提供全方面数据支持。



为什么选择 CDA？

“Certified Data Analyst Institute”与经管之家（原人大经济论坛）成员是来自学界、实务界，国内大陆、台湾及国外数据分析和数据挖掘相关领域的教授、专家、工程师及企业高端人才，名师荟萃，学术浓厚，技术前沿，代表了国内数据分析研究领域的专业水平。CDA 数据分析师等级标准是根据过内各大企业对人才技术的需求而设立旨在为国内数据分析发展阶段提供一个全面、科学、专业的标准规范，说明究竟什么人才是优秀的数据分析师？

CDA 数据分析师的就业前景可选择于通讯、医疗、银行、证券、保险、制造、商业、市场研究、科研、教育等多个行业和领域。，根据三个不同的等级胜任不同的数据分析工作任务。

如何成为一名 CDA 数据分析师？

经管之家（原人大经济论坛）为 CDA 数据分析师唯一考试报名中心，CDA 数据分析师 LEVEL ⅠⅡ系统培训，由经管之家根据 CDA 认证标准而设立的一套针对数据分析师技能的全面系统培训。培训师资目前均来自学界、实务界相关领域的讲师、教授、专家、工程师以及企业资深分析师，名师荟萃，代表了国内数据分析培训的专业水平，可以很好地保证培训的学员既能学到扎实的数据分析理论知识，又能具备较强的利用软件解决实际问题的能力，保证学员能胜任各行业数据分析师工作的要求。

学员可以参加“CDA 数据分析师 LEVEL Ⅰ系统培训”或”CDA 数据分析师 LEVEL Ⅱ系统培训“进行深入学习。学习后可报考 CDA 数据分析师等级资格考试（点击**考试详情**），通过考试后即可获得 CDA 数据分析师等级资格认证证书，后续可参加 CDA 线下聚会、沙龙、招聘会、研讨会等各项活动，促进数据分析师领域的高效发展。

经管之家（原人大经济论坛）CDA 优势：

人大经济论坛（bbs.pinggu.org）：于 2003 年成立与人大经济学院，致力于推动经济学科的进步，传播优秀教育资源，目前已经发展成为国内大型的经济、管理、金融、统计类的在线教育和咨询网站，也是国内最活跃和具影响力的经管类网络社区。

人大经济论坛运营团队：北京国富如荷网络科技有限公司，成立于 2006 年 6 月，公司以人大经济论坛为运营平台，经营业务包括培训业务、数据处理和分析服务和教辅产品等。人大经济论坛“数据分析培训中心”自 2007 年成立以来，致力于开展统计软件、数据分析和数据挖掘的培训与咨询服务，目前已有专家、讲师团队 100 多位，拥有自主版权的视频课程 60 多门，每年开设现场培训班 100 余场，建立了完备的数据分析培训课程体系，每年培训学员 3000 多人。服务过的企业包括中国电子商务中心、招商银行、中国人民银行、中国邮政储蓄、中国联通、中国汽车技术研究中心、南京梅钢等机构。

发展至今，人大经济论坛 “数据分析培训中心” 已经成为具有影响力和知名度的数据分析培训机构，我们一直努力做到：将数据分析变成一门常识，让统计软件成为学术研究的好伙伴，企业经营的好军师。

优势一：师资

CDA 数据分析师系统培训，由经管之家根据 CDA 认证标准而设立的一套针对数据分析师技能的全面系统培训。培训师资目前均来自学界、实务界相关领域的讲师、教授、专家、工程师以及企业资深分析师，名师荟萃，代表了国内数据分析培训的专业水平，可以很好地保证培训的学员既能学到扎实的数据分析理论知识，又能具备较强的利用软件解决实际问题的能力，保证学员能胜任各行业数据分析师工作的要求。CDA 数据分析师培训注重结合实际，把具技术含量、具价值理念的课程传授给学员。课程还注重启发式教学，让学员在动手解决问题中去学习。

CDA 数据分析师课程的大纲和内容，既由经管之家和国内数据分析、数据挖掘领域专家潜心开发和反复研究，又经过科学的调研确定，并且将不断地随着数据分析的市场需求和数据分析技术的发展而调整，课程内容始终关注市场、关注前沿。课程内容的设计更注重阶梯化、体系化的原则，每一个学员，不论学习和工作的背景如何，都能在该课程体系中很快找到适合自己的课程，并不断学习提高。

优势二：继续学习

所有 CDA 学员除了学习现场课程之外，还会得到全程视频录像及辅助学习视频课程（包括 SPSS,SAS，R 语言及数据挖掘等内容），此系列视频课程可以进行后期巩固学习和进修学习，可扎实现学技能、拓展课余知识、升华技术层级。

优势三：在线学习

Peixun.net（经管之家在线课堂）学习视频可实现随时随地在线听课，10 分钟一小节，可进行个性化、碎片化学习，更具针对性与便利性。CDA 上课方式分为现场及远程两种方式，远程在线学习引进了最新设备与技术，解决了各地区学员的需求，并实现了如同现场般的远程答疑及讨论。

优势三：交流社区

经管之家有十二个社区，七百个版块，五百多万会员。每日讨论的热点话题及资料以千计。学员在学后可以到”CDA 数据分析师”版块进行交流、提问、下载资料等，形成数据分析专业聚集地，促进学员在圈子交流中高效发展。

CDA 协会网址：www.data-analysts.org

CDA 微博互动：@CDA 数据分析师_人大经济论坛

CDA 微信订阅：CDA 数据分析师





读后感和推介

附录

其它版主/坛友的读后感和推介

〖**夏目贵志版主：**这本书向为数众多的以使用计量经济学为目的学生和学者们展示了一位有能力、有勇气、有恒心的学者从入门计量经济的学习到应用计量经济方法的经历、经验、心得。同时，该书也系统的解答了大量关于计量知识和数据处理方面实际中最常见的问题，对于本科及研究生阶段，尤其是刚开始系统学习计量经济学的学生和学者们，是非常有益的补充。〗

尾声 1：

本书成书期间 73 位网友的书评·回贴·反馈

注：这本书的原始内容，始发于经管之家（原人大经济论坛）。相见时难别亦难，在本书即将和您说再见的时候，我们将本书在成书和初期撰写期间，广大网友对本书的关注. 评论和反馈整理出来，感谢一路关注本书的所有经管人. 财经人。

1. prescottwong 坛友：

很有启发，受益匪浅哈。对于计量，想说爱你不容易，我自己学计量也是自己一点一点抠下来，全靠自己努力，学校老师很少系统教授，有些他们也不懂。所以，国内的小硕们还得自己静下心去坑书本，玩命的一遍又一遍在夜深人静的时候跑数据。我的硕士论文就是半个月整理数据，建模型，三个通宵在跑数据. 修正模型，然后又是绞尽脑汁写分析些对策，一个月体重直接下降 6 斤。再回首那段岁月自己能坚持下来确实不容易。自学计量确实是一个下苦功的活儿，不过年轻就应该拼一拼，不拼哪知道自己的潜能尼。

2. 明阳^_^坛友：

能看到这么优秀的帖子，这个中午实在是赚了

3. 孤单的我们坛友：

感谢分享！学习的过程的中，把理论与实践结合起来，会更好的理解不同模型的区别与适应条件。 孤军奋战固然需要，交流与培训则是更上一层楼。

4. C 程序语言坛友：

谢谢楼主的分享,能够看到这么有营养的帖子,感觉很幸运,也祝楼主好人有好报,一直顺利下去~~

5. wangziguilai 坛友：

最近在上计量，曾经很迷信计量是万能的，后来发现，他只是一个工具，重要的是自己的逻辑思维，自己要知道建模是为了什么。

6. springrainq 坛友：

求人不如求己，, 非常赞同。我写博士论文的时候，用时间序列数据，通过模型结构突变来判断某个经济指标发生突变的时间以及引起突变的因素。没有经验可借鉴，国内论文没有一篇完整使用该方法的，抓狂，最后还是啃外文资料一步步推进，苦不堪言，但收获颇丰！共勉！

7. YY.HH 坛友：

谢谢楼主！好棒！文科生要读研了，也想自学计量经济学，但很怕，看完很

受鼓舞！

8. 断魄书生坛友：

谢谢加分。共勉！看了你的两篇文章，对后来人相当有帮助。其实昨天想说把你的两篇文章发表到我的一个经济学 公众号 上的，但是昨天太忙了，没来得及说完。 我的公众号"Avalon 应用经济学”，非盈利，师友圈子，主要致力于经济学学习和应用。求授权

9. cliocao 坛友：

像做交易一样，如果只是提方法论（罗列一些列的书名）而不提心境和心态，虽然也能引起重视，但是终究不适用。而楼主更多的给出了一些主观的看法和心境，值得借鉴，感谢！

10.iragh 坛友：

楼主的成功经历，真诚地态度，详尽的描述，让我这个外行对学习计量经济学增加了不少信心。 谢谢楼主！

11. karst 坛友：

我现在的专业，也是跨专业学习的。深知其中的苦处。向 LZ 表达敬意。也从中学到了一些有益的东西。对以后的学习有好处。因为知识的交叉性和边缘性趋势越来越明显。学习未知知识的问题将越来越重要。

12.admaijh 坛友：

只要有恒心，多琢磨，借助人大经济论坛这个平台，掌握计量经济学是可以的。

13.Xiavi 坛友：

我是台灣來的，偶然來到人大拜讀樓主的文章受益匪淺! 真心謝謝樓主寫出自己學習計量的心路歷程，期待很多好帖!! PS. 祝中秋節快樂~

14. muyuxiu5623 坛友：

真的很谢谢您，做学问就需要像您这样，榜样，以此激励自己前行。

15. xyq172100 坛友：

楼主在上，请受小弟一拜！ 任何学科的学习都是由浅入深步步为营，楼主的亲身经历充分地证明了这一点。 可贵的是，楼主把自己在学习过程中的心态变化细致地重现出来，并且站在事后的角度为大家剖析，有理有据，不仅让我觉得身临其境，还特别有说服力，令人印象深刻。 最重要的，不在于楼主对于知识的掌握，而在于字里行间渗透出的无私和实事求是的态度，这更些值得我不断追求。谢谢！

16. 竹林听水坛友：

今天我导师专门在课堂上提起楼主的这篇文章了，才发现原来楼主是我导师的学弟向您学习

17. shihua 坛友：

虽然楼主您说自己是外行，但看的懂那么多高深的书，我很佩服，觉得你也一定是数学很好的了。我本是文科生，原也想认真学习计量，买来高铁梅老师的书，一看很多公式就直接晕菜了。但愿从你的经验中能够找到合适自己的路。羡慕数学好的童鞋~~

18. pzh_hzp 坛友：

求人不如求己，很多东西是需要自己去琢磨的。要想取得成果，非下苦工不可。]好,赞一个

19. fppyjw 坛友：

正在学习中，和楼主类似，本科就是些数理统计的皮毛，现在工作需要重新捡起来，需要系统的统计和计量分析，初期还是感觉压力很大，怕自己拿不下来，楼主现身说法让我很受鼓舞，还是要系统的学习，一步步一遍遍来才扎实，! 另楼主能不建个计量学习群，给我们这些刚开始的人一个交流的地方，谢谢

20. akuang20 坛友：

有心才有行动，所以终身受用~

21. pat 乐坛友：

太感动了，把门外汉和半吊子描述的如此贴切而没有鄙夷之色，知之为知之，夯实基础，共勉之

22. VernonLin 坛友：

支持啊~ 完整看了 2 遍，LZ 发这么多好帖子，果然是热心的好版主。 收藏了，相信以后每次回来看都有不一样的收获

23. 夏目贵志坛友：

说的很好啊！还有一个点可能是主要通过自学难以实现的，那就是实践。其实绝大多数论文明里只说了作者所做的工作的冰山一角，尤其是应用计量论文。模型的选择和这个选择的背后的故事谁也不会说出来的。这就给以自学为主的人带来了很大的困难。计量理论虽然难，但是毕竟是数学和统计学，对就是对，错就是错。任何人都能靠自己的力量学懂。而没有实际经验的人（或者实际经验不足）的人，是很难知道书本上的东西实际中是如何使用的。这个看再多的书也没用。必须得有老师手把手的教，一点一点的讲解。所以楼主说的很对的一点就是有可能的话还是要接受专业的训练。

24. tina_wu 坛友：

真心被楼主孜孜不倦追求学术的精神折服，佩服，学习！顿感人跟人的差距就是在这每天的努力与荒废中日积月累的，追忆自己逝去的时光不免后背发凉啊

25. xixia333 坛友：

楼主的学习历程很值得肯定，因为从你的身上传达了一种勤奋的正能量。很多人在学习的过程中都会有或多或少的恐惧心理，总怀疑自己的能力，或者也有可能是懒惰吧。你的经历告诉我们，做任何事需要一种坚持，做自己的事，思考

自己的问题，真正花时间，真正用心。人需要终生学习，仅靠已有的知识是走不远的。我现在是厦大计量博一的学生，工作三年后辞职考的博士，目前计量丢的时间比较长，学习感觉有点吃力，但我有信心学好，勤能补拙是良训，以后向你多学习交流。感谢论坛！

26. a649376826 坛友：

谢谢楼主分享经验，受益匪浅，给我们这些门外汉指明了道路。

27. zhoujielunli 坛友：

三读作者的文章，身为您的学术态度和热情所感染，自己马上就要开学了，没有学习，反倒是不断的思索过去的经历，对于学术态度的不端正，没有兴趣的指引全是功利性的行动，之前的硕士经历已经弊端种种，希望自己能够真的做到端正态度，珍惜来之不易的学习机会，九月十号上财见，谢谢楼主，以后会一直关注楼主的文章，学习对学术的态度，以及对学习的热情。

28. 十年树木坛友：

有时候专门为某一个问题去学习一种新方法效率会很高，如果自己漫无目的的去学习，发现效率很低，学着后面的前面的忘记了，等学完后基本上都忘记差不多了，将来用的时候还得从新扒资料学习。 当然这种学习的缺点，正如兄弟所言是一种零碎的知识，不系统。优点是有针对性，效率相对高些。

29.statax 坛友：

好文！业界良心啊。写得太长，快速看了一下，楼主很用心。 我的观点是，在有了一定基础后，学什么方法只要找到资料，还是容易掌握的，所以，基础很重要，象楼主说的伍里奇的书就很适合打基础。技术故然重要，看文献和写作是看计量无法替代的，有了基础，陈强和卡梅隆的 stata 教材作为 cookbook，再看一些专著，象空间计量，HLM，分位数，时间序列之类的，大可不必花太多时间，需要时拿来用即可。多写多看多发表才是王道。。。

30. lwzxy 坛友：

热心. 实在，楼主的一贯风格也。

31. niuniuyiwan 坛友：

感谢师兄分享自己的心路历程，为伊消得人憔悴，衣带渐宽终不悔，祝一切都好!

32. paulobbc 坛友：

不能够更同意你的看法了，还是要静心学习理论知识。谢谢分享。

33. qinnanfreedom 坛友：

写得真好，确实是真实经历过的心得！!！赞一个

34. yulinwei 坛友：

楼主实在厉害，也特别用心，令人敬佩

35. Jonerice 坛友：

徐兄的观点很是精辟。关于计量经济学的学习和应用，我(ˇ_ˇ）想～除了“海龟”能够在国外学校的课堂上系统的学习计量，而在国内的广大“土鳖”只能是“师傅领进门，修行在个人”式的学习计量。徐兄的经历也是广大计量人的的经历，我的计量也是大部分靠自己学的。这种“痛，并快乐着”的心路历程每个计量人都会深受感受。所以，很有幸能够在人大经济论坛认识志同道合的计量朋友，与之交流，或学习，或吸取营养，或受到启发，从此“不再是一个人在战斗”。谢谢。

36. hanyusi 坛友：

用心的好文章 学术精神！ 谢谢并祝福楼主！

37. briantding 坛友：

感谢楼主分享。读文后收获：1. 必须是好酒。2. 是好酒，还要突出咱家酒得特点。3. 证明好酒的过程不容易。

38. EI_charlie 坛友：

非常喜欢这样用心的分享，

39. hexiangwu 坛友：

前辈的经验介绍的很详细到位，很受用！多谢

40.最爱拜仁 坛友：

看了楼主的文章，真的感觉学到很多，真的要好好学下英语写作了

41.changqingmingzi 坛友：

虽然我不是搞计量经济学的，但是看到楼主的心得也很有搜获

42. zhoujielunli 坛友：

赞赞，九月份要开学了读博了，看到您的文章很有启迪啊，最近自己在用 stata 跑伍德里奇的导论还有三章跑完了，整个过程很痛苦，但是看到您的文章，自己觉得这个过程是必须的，您的之前的学习计量的帖子，我也看了，非常共鸣，您的留言给了鼓励，谢谢，好文必须要赞，希望楼主多谢好文。~~~~

43.诸葛山下坛友：

楼主写的真好，有“道”有“术”，解决了一些我的困惑。虽然和我专业不一样，但是道理都一样，很有借鉴意义。我们老师常说，读研究生主要锻炼了逻辑思考和写作能力，此种说法在文中有验证。

44. 飞翔的小小狗 坛友：

真真的好帖子，正准备写论文，受教了，赞一个~~

45. ruoshui2015 坛友：

感谢楼主分享投稿经验，我是认认真真学习了，受益匪浅，投稿刚刚被拒。

46. jay6391268sxa 坛友：

经验很好 可以避免后来者走弯路 谢谢分享 祝越来越好

47. 琥珀糖坛友

好久没看到这种好文了，顶

48. Crsky7 坛友：

好文章好分享好心得好经典

49.ryjsunny 坛友：

之前投了第一篇文章不到三个月就被拒了，后来投了个影响因子低点的，将近半年给了修改意见，最近在苦逼的修改中，今天刚好看到你的帖子，也顺便浏览下你的其他帖子，我比较关心的是，在你的这个帖子中：

〖"万万"个论坛币〗计量实证的“摸底”考试

你的问题总结的却是很实用，就像一个手册一样，从计量研究着手，从对数据判断，到数据描述以及之后的技术处理，我觉得对经济实证研究非常有帮助。就是想知道，根据这些问题你有没有把相应的答案，或者说相关的讨论总结下(应该有些问题现在也是在讨论中，没有标准的对与错）可以的话，可以跟大家分享下你最后总结的结果！这将是一个伟大的工程，很期待，或者你如果有的话可以发我一份吗？

50. jjianmin 坛友：

学习了。虽然离发 SCI 等距离还比较遥远，但是很多经验真是共通的，个人在国内 CSSCI 等投稿也是碰到类似情况。所以很有借鉴意义

51. bling1987 坛友：

看了楼主分享的心路历程，深有同感啊，写 SCI 论文真不容易，期间各种摸索和心酸啊，我是经济学跨到地理学，历经各种苦逼煎熬写了好几篇成品和半成品，直到博三才发表了一篇 SCI 刚刚达到毕业要求，博士都不容易啊

52. 骑猪的麦子坛友：

难得看到最佳人气板块出来的高质量文章 赞一个！!

53. 征夷大将军坛友：

学长，您第一次发帖我就在关注您，我也是搞农经的，农经学子的最大优势在与农科与农经都吃得开。您果然有大才，这篇介绍发表英文文章的帖子写的非常详细，很有价值，对我很有帮助，燃起了我投 SSCI 的欲望，我要向您多学习。Publish or perish 是我一直信奉的观点，祝您妙笔生花，发表更多高质量的论文。

54. betterljh 坛友：

楼主的经历和我太像了，不同在于我是 eviews 起步的，懂得了 eivews 的基本操作，但是深入分析还不够，现在准备先从计量的一些基础理论入手。正在看清华大学出版社 孙敬水的书，感觉还不错，部分看不懂，有一些原理使我懂得更多了。伍德里奇的书买了，打算下一本看它。

55. 溪飞白鸟坛友：

非常有用。以楼主为镜，估计能少走很多弯路

56. 古月 Tina 坛友：

赞一个，楼主好牛啊，竟然啃下了这么多难啃的书，我们现在也在学习各种统计分析软件，计量经济学是我们的专业课，每次我想认真学习的时候看着看着总是望而却步，楼主给了我很好的激励……谢谢楼主的经验分享

57. jiangqzox 坛友：

这个写得真好,真实地反映了学以致用,用中再学的良性发展互动的情景

58. apoda 坛友：

经验非常贴合实际，虽然我还在学习的路上，但是看了很有共鸣，实在是难得遇见的好帖。 满满的都是正能量和指路牌，谢谢分享！

59.renjiang_yd 坛友：

很宝贵的切身体会，对我有很大的激励作用，谢谢了！

60.flying_08 坛友：

看起来确是切身体会，于娓娓述说之中介绍了很多宝贵的感悟和经验。学习了，多谢分享！!

61. lnulxg 坛友：

认真地走过，才知道什么是辛苦和收获！值得赞！

62. fred_7412 坛友：

真的写得很好很真切！对于计量学了中级，但是仍然懵懵懂懂的人有不少的

指导，继续努力，向前辈靠近！

63. annyding345 坛友：

楼主，感觉你写的很好。我现在也是刚看完古扎拉蒂的《计量经济学基础》，但是对于啥时候取对数，怎样解释还是模糊的，伍德里奇的《计量经济学导论》也买了，一直在书柜了睡大觉，惭愧啊！

64. 才才科坛友：

也是一个外行，刚踏上计量之路，谢谢你的指引。

65. Jealy 坛友：

我大学学习的金融和理财，超级喜欢计量经济学，可是大学老师说计量经济学太难了，没有给我们开设这门课程，感觉好遗憾。大学期间的计量经济学基本都是靠自学的，看的是西南财经大学庞浩老师的课程。希望和楼主多交流沟通

66. 落陷的阳光坛友：

满满的正能量，谢谢分享！现在正在学计量经济学，为自己的不认真感到惭愧！

67. Spacelin09 坛友：

那时不知道要对数据做基本的处理，如处理极端异常值，缺失值. 对一些变量进行描述性统计分析，做散点图看趋势等等，只知道将自变量和因变量分别统计出来，然后就建模了”还真是超级吻合，我现在就是出于这样一种状况哎

68. 敏钥儿 1990 坛友：

很欣赏楼主的学习精神，也确实如此，学习都是不断摸索的过程，没有摸索，便没有进步，为楼主的精神赞一个！

69. fuzzylog 坛友：

作为一个同样曾经学过计量的外行人来说，很有感触。

70. bfzldh 坛友：

昨夜西风凋碧树，独上高楼，望断天涯路 衣带渐宽终不悔，为伊消得人憔悴众里寻他千百度，蓦然回首，那人却在灯火阑珊处

71. 丁小文坛友：

金融研究生，上到研二还觉得啥也不会，深深佩服楼主的毅力，心向往之

72. 独眼鱼坛友：

感谢楼主分享的计量学习的经验和感悟，其中的很多东西都十分受用，谢谢

73. fdhdd 坛友

这种自己心得的分享很难得,赞楼主

尾声 2 ：

本书第一个宣传专题发布时 43 位网友的书评·反馈·热议

本书第一个宣传主题请见：

统计学还是虐待学？由你决定！论坛为自家坛友出书啦！

注：2015 年 12 月 21 日，本书的第一个宣传专题贴在经管之家正式发布，短短 6 天时间，超过 12000 名网友关注了本书，并有超过 540 个热烈回复，贴子的主楼码的非常之高，队形也非常之长，感谢大家的冬日热情和对我们的鼓励。在本书的告别之际，我们将这些宝贵的有代表性的评论和意见整理出来，感谢所有关注本书的朋友们，这本书是一场对知识和求知无始无终的穿越，不说再见……

1. 夏目贵志版主的意见和反馈（ps:非常感谢他在本书成书期间的支持，他的反馈意见已经被主创团队认真收录，并将在本书第二版中改进和完善）：

这本书向为数众多的以使用计量经济学为目的学生和学者们展示了一位有能力、有勇气、有恒心的学者从入门计量经济的学习到应用计量经济方法的经历、经验、心得。同时，该书也系统的解答了大量关于计量知识和数据处理方面实际

中最常见的问题，对于本科及研究生阶段，尤其是刚开始系统学习计量经济学的学生和学者们，是非常有益的补充。

首先，我觉得编写这本书是一个非常有意义的尝试。就我个人有限的了解，长期以来，计量经济学的教学就以理论知识为主。学生对实际问题的思维和应用能力一般都被认为是天赋，而非努力培养的结果。院校招生的时候要“有学术研究潜质”的学生，一部分指的就是这个，即，在掌握理论知识之后，有创造性的应用理论知识的能力。遗憾的是，很多人，即使是顶级名校毕业，在这个方面也不是非常成功。在这个背景下，这本书能够以独立于理论书籍却又不脱离理论的方式，探讨一些学习方法和心得，从早期开始培养学生，强调思维方式和学习态度，应该会对很多学生有所帮助。当然，受限于书中资料的来源和书的篇幅，书中各个部分独立性比较强。这既是优点也是缺点。说是优点是因为这种结构允许读者自取所需，而不局限于某一种思维和学习定式（和课堂的线性教学方式不同）。说是缺点是因为与此同时，书中的内容在结构上并不紧凑，即，各部分之间的联系并不明显。怎么在有序和灵活两者之间取舍，就看编者如何改进了

2. analysismath 坛友：

作为一个计量的门外汉，还能不被作者鄙视，感激涕零！ 希望得到此书，可以不被其他人鄙视。

3. xuguw 坛友：

计量统计学习亲历与经验分享. 经典而热门的疑难问题答疑解惑与详解……很棒的！！ 地球人都想看的！！有效回复，对本书示忠示诚示爱， 努力取得预订权，速得电子书！！

4. tianweimin2962 坛友：

免费领书，我想要书！ 我真的很想要这本书啊，这是真正的经验之谈！

5. 鹤归孤山坛友：

我想要书 ,免费领书。作为一只还没入门的菜鸟很想看到点适合入门的东西！

6. bnbth 坛友：

我想要书，统计学毕业的孩纸，非常期待该书，谢谢

7. Xiavi 坛友：

聽起來好棒的書，我也想領書!!! 另外，不知道那些通路買的到這本書呢? 人在海外希望能知道詳細購買方式~ 恭賀出書！

8. walqy222 坛友：

我想要书，感觉挺适合我的，计量小白从头开始

9. zhangpeng1975 坛友：

我真的需要这样的好书，感激！

10.旧时光是个美人坛友：

我想要书 可以没有币 一定要有书啊！！ 如果有纸质版的更好了！！

11. sxufelyg 坛友：

论坛里的楼主们，辛苦了一年，这是你们的收获，催人奋进，值得学习。光看目录就知道书相当不错。谢谢

12.appearben 坛友：

我想要书，免费赠书。 以后想做研究，这本书太有帮助，希望有这本书可以看！!!

13.andalis 坛友：

我想要书 自学计量统计还是比较痛苦的（本人主攻政治经济学），耗费大量的精力

14.醴陵人坛友：

统计系毕业的孩纸，工作后难找到这么好的资料，跪求，谢谢！

15.ksr 坛友：

作为金融民工，统计知识和软件应用是绕不开的，坛主的好书一定要支持！!我想要书！!!

16. piaoyingyixiao 坛友：

所有经济学的基础，就是学会数据分析。 不管怎么样，先把书抢到看看

17. 瘋樂坛友：

免费领书！我想要书！强烈需求，自己看计量经济真是很迷糊，谢谢论坛能

有这么一本答疑解惑的书，这真是极好的！

18. 烟火人间坛友：

我想要书~学霸起来~论文 fighting~

19. FLD 坛友：

我想要书，经济学专业出身，什么都不精通，现在大数据的概念被炒得越来越热，刚毕业，在公司的价值估计也就是数据分析上的一点点，但仍然是可要可不要的那种，现在只能小心翼翼，静下心来，沉淀，问了有朝一日可以一鸣惊人

20. afu_ty 坛友：

虽然接触统计有一段时间了，但对一些关键的东西还是云里雾里，如果有这么一本通俗易懂言简意赅的神书，那想必是极好的。这样一来再也不用担心自己学不会统计与计量了。想想面对无从下手的数据，一直梦想着有人指点迷津，有书来醍醐灌顶......苦苦的等待之后，你终于来了......我要学好统计计量，我要一本好书！

21. Captain-CUI 乐坛友：

我想要书，痛苦的论文马上就要来了，计量还没有准备好，这可怎么办？？？

22. xiyubaobei0213 坛友：

我想要书，免费领书。太喜欢了，我学的是统计学，现在工作了，还是处处与统计学息息相关，好书就要抢先读~

23. 爱看你微笑坛友：

我想要书，免费领书。多学点东西总是好的，技多不压身。学习……

24. yjhist 坛友：

本人学气象，常常用来处理大量的数据统计，工作后才发现数学里概率统计学科的重要。以前随来潜水，这次得冒个泡请求给留本书吧，不同行业，相同目标，交流交流。谢谢版主！

25. 万里晴空坛友：

看来是本好书啊，什么时候全部出齐呢？问题意识为上，工具方法为下。

26. 爱发呆的小丸子坛友：

中文狗学统计学，学 stata，要多虐心就有多虐心。神啊，赐我秘籍一本，看懂图标，懂得解释，驱赶数据恐惧心理，各种大招！!!

27. 西门城坛友：

作为一名菜鸟 很希望看这样的书

28. chenshude1989 坛友：

师兄说的太好了，本人刚开始学计量的时候，只是为了会用，根本就没有过多的去理解计量的基本思想，经过一段时间之后，对计量感觉还是很空洞，感觉自己没真正学到东西，所以我现在还是从最简单的计量开始学起，掌握计量的思想，才能更好地进行实证研究，很希望能得到这本书，进行深入的学习

29.stanpeter 坛友：

特别特别想要书，你不给我书就是不和我一起建设社会主义和谐社会。 我要领书~~~

30. 小熊笨笨坛友：

对计量. 统计很感兴趣！ 我想要书 ！ 谢谢！ ps 这算三行情书嘛 哈哈

31. 20091548 坛友：

我想要书，虽是雾霾天，可是药不能停，学习之路不能停！!!

32. tfhis 坛友：

有大神为我们领路，真是幸福。期待这部电子书

33. xuxinqiujiao 坛友：

免费领书 正准备学习计量经济学，最好是少花费很多无谓的摸索时间

34. johnvone 坛友：

已经工作，还是喜欢研究计量. 量化的知识，我想要书。

35. zbingchao 坛友：

哎，一直被统计和计量所折磨，导师是数量经济学出身，我是管理出身，跟的好辛苦，统计计量的各种教材十几本了，人大的统计学系列教材八本，吴喜之的统计学，还有什么爱上统计学，深入浅出统计学，李子奈的计量经济学，伍德里奇的计量经济学导论，计量经济学基础，赵国庆的计量经济学入门，以及各种

stata 技能的学习书，也七八本，汉密尔顿，stata 手册，胡博 stata 分析，人大刘传波的十八讲，以及各种资料，真心扛不住了，希望有本神书来解脱我，期待这本书~！如果出纸质书，我一定多买几本，送我师门

36. xuxiangyang7 坛友：

我想要书，计量虐我万千遍，我还是不能和她分手，不能和她分手。。。

37. amimadoka 坛友：

经常在论坛中学习，这次的书更是让人期待。希望有更多的好书陆续出来。谢谢这些牛人们的贡献。

38. yijiaobani 坛友：

顶起来，中国需要这样的书，把统计拉下来。

39. seasrlife 坛友：

每天都来看看这本大神们的巨作是否已发出，焦灼的心期待着，盼望着……对于一个没有统计学基础的管理学 Doctor，这可是救命的神药啊！ 我想要书！我想要书！我想要书！重要的事情说三遍！ 我想要书！我想要书！我想要书！重要的事情再说三遍！ 早点出炉吧，我要学习. 学习. 在学习！

40. 独一坛友：

想要免费的书，更想借鉴作者的学习经验。谢谢赠予。

41. comcwal 坛友：

当初统计学学的一塌糊涂 现在终于可以补补了 我想要书！!

42. szm0818 坛友：

我很想要书，免费领书，谢谢。 论坛越来越有内容了，希望越办越好!

43. ━┈_D 大調° 坛友：

论坛新人~ 值此论坛神书之际~特发此贴~以表想要此电子书的决心~ 我想要书~ 我要此书~ 我要电子书~ 免费领书~ 砸我！

本书涉及的120个计量统计经典答疑索清单表

一、软件学习

1.1 学习方法

问 题	链 接
求助：实证研究零基础博士生请教怎样选择、运用模型进行数据分析，看什么书可以入门？	http://bbs.pinggu.org/thread-3782579-1-1.html
想学习数据分析，毫无基础，应该从哪些软件入手	http://bbs.pinggu.org/thread-3793702-1-1.html
正在看张文彤的《SPSS统计分析高级教程》，看不懂怎么破？	http://bbs.pinggu.org/thread-3748197-1-1.html
SPSS在数学建模里有什么主要的作用？	http://bbs.pinggu.org/thread-3788473-1-1.html

1.2 SPSS

问 题	链 接

应该如何对连续变量进行设置？	http://bbs.pinggu.org/thread-2633633-1-1.html
请问 SPSS 中导入的数据中默认都是字符串，改成数值型	http://bbs.pinggu.org/thread-3755176-1-1.html
怎么使用计算变量分解多选项问题.挺急的,在线等!	http://bbs.pinggu.org/thread-3757193-1-1.html
SPSS 中计算变量生产新变量的还原问题	http://bbs.pinggu.org/thread-3764404-1-1.html
SPSS 怎样在日期变量的原始数据中取年度值？	http://bbs.pinggu.org/thread-3761449-1-1.html
怎样将一个变量分为两个？？？	http://bbs.pinggu.org/thread-3783628-1-1.html
求推荐从建模到结论的书籍 附带SPSS或SAS程序。	http://bbs.pinggu.org/thread-3754365-1-1.html
初学者怎么学习 AMOS 软件啊？	http://bbs.pinggu.org/thread-3753792-1-1.html
关于 SPSS 自由度问题	http://bbs.pinggu.org/thread-3761984-1-1.html
统计报表结果用脚本导出的疑问	http://bbs.pinggu.org/thread-3760692-1-1.html
急！! SPSS 回归结果中 星号标记	http://bbs.pinggu.org/thread-3749109-1-1.html

1.3 STATA

问 题	链 接
STATA 跑数据	http://bbs.pinggu.org/thread-3770340-1-1.html

用 STATA 做的图，麻烦大神给指点一下	http://bbs.pinggu.org/thread-3779337-1-1.html
【求助】STATA 怎么保存啊	http://bbs.pinggu.org/thread-3790491-1-1.html
STATA 统计性描述	http://bbs.pinggu.org/thread-3783429-1-1.html
表格在 STATA 里太长了怎么破	http://bbs.pinggu.org/thread-3772156-1-1.html
STATA 中缺省值处理	http://bbs.pinggu.org/thread-3783008-1-1.html
【求问】STATA 一介差分的标准差怎么做啊	http://bbs.pinggu.org/thread-3754093-1-1.html
STATA 双击变量会在 command 里出现变量名，那么 do 文件里怎么才可以快速出现变量名？	http://bbs.pinggu.org/thread-3785169-1-1.html
有没有特别全面的 STATA 教材？麻烦大家推荐一本教材。	http://bbs.pinggu.org/thread-3775742-1-1.html
想学习 STATA，但不知从何开始	http://bbs.pinggu.org/thread-3790477-1-1.html

二、多元分析

2.1 联合分析&聚类

问 题	链 接
新手-联合分析问题请教大神们	http://bbs.pinggu.org/thread-3760590-1-1.html
SPSS 聚类分析 类间距离如何获得？	http://bbs.pinggu.org/thread-3763944-1-1.html

求助 SPSS 对取值为 0、1 的不同分类变量做聚类分析？	http://bbs.pinggu.org/thread-3774323-1-1.html

2.2 主成分&因子

问 题	链 接
SPSS 做因子分析，最后发现因子综合得分均值为零，什么原因？	http://bbs.pinggu.org/thread-3744403-1-1.html
SPSS 因子分析后的 anova	http://bbs.pinggu.org/thread-3752490-1-1.html
SPSS 怎么用主成分分析后的变量和因变量做一元线性回归	http://bbs.pinggu.org/thread-3746720-1-1.html
有关主成分分析	http://bbs.pinggu.org/thread-3757786-1-1.html
STATA 主成分分析	http://bbs.pinggu.org/thread-3771766-1-1.html
关于主成分分析中变量的一个问题	http://bbs.pinggu.org/thread-3780889-1-1.html
对提取的主成份进行线性回归	http://bbs.pinggu.org/thread-3785872-1-1.html
能不能用因子分析和回归分析做宏观经济分析	http://bbs.pinggu.org/thread-3792426-1-1.html
怎么解释主成分分析的得分图？	http://bbs.pinggu.org/thread-3795483-1-1.html

三、回归

3.1 简单回归

1、变量问题

问 题	链 接
求大神！SPSS 回归分析中如何将自变量按时间分类进入模型？	http://bbs.pinggu.org/thread-3787377-1-1.html
[求助] 模型消除异方差后，如何引入虚拟变量	http://bbs.pinggu.org/thread-3756278-1-1.html
遗漏变量问题！	http://bbs.pinggu.org/thread-3751322-1-1.html
哑变量结果解读	http://bbs.pinggu.org/thread-3796440-1-1.html
同时存在两个虚拟变量的模型，哪个为默认的基准组？	http://bbs.pinggu.org/thread-3771084-1-1.html
混合线性模型的因子能不能是连续变量？	http://bbs.pinggu.org/thread-3752078-1-1.html
菜鸟求问找对照组和不找对照组有什么区别	http://bbs.pinggu.org/thread-3790279-1-1.html
求大神，关于异常值的处理	http://bbs.pinggu.org/thread-3750393-1-1.html

2、交互、中介、调节效应

问 题	链 接
SPSS 线性回归的调节变量	http://bbs.pinggu.org/thread-3789523-1-1.html

关于倒 U 型的调节效应	http://bbs.pinggu.org/thread-3759376-1-1.html
交叉项中存在对数形式，如何解释？	http://ask.pinggu.org/?q-557.html
调节变量是连续变量，自变量是分类变量，交互效应需要中心化吗	http://bbs.pinggu.org/thread-3063467-1-1.html
[疑难问答] 我想问调节效应是正的，但是调节变量是负的，可以吗	http://bbs.pinggu.org/thread-2714854-1-1.html
[学科前沿] 如何解释交互作用	http://bbs.pinggu.org/thread-3431500-1-1.html
交互作用	http://bbs.pinggu.org/thread-3915409-1-1.html
虚拟变量不同基组交叉项系数的显著性	http://bbs.pinggu.org/thread-3876111-1-1.html

3、OLS（最小二乘法）回归、逐步回归、泊松回归、稳健回归

问 题	链 接
判断 Enter 方式和 Forward Stepwise(Conditional)方式两个模型哪个	http://bbs.pinggu.org/thread-3786690-1-1.html
求助：关于计量方程回归结果比较大小问题	http://bbs.pinggu.org/thread-3763823-1-1.html
谁知道哪本书里面有具体讲泊松回归负二项回归以及弹性分析的内容	http://bbs.pinggu.org/thread-3717839-1-1.html

怎么做 robust 的 ols 呀？	http://bbs.pinggu.org/thread-3762928-1-1.html
关于概率的问题	http://bbs.pinggu.org/thread-3770253-1-1.html

3.2 分类数据回归

问 题	链 接
logit 回归如何计算边际效应？用 SPSS 软件可以计算出来吗	http://bbs.pinggu.org/thread-3752076-1-1.html
logistic 回归如何将频数做为因变量	http://bbs.pinggu.org/thread-3754031-1-1.html
多分类无序标量	http://bbs.pinggu.org/thread-3752686-1-1.html
logistic 回归	http://bbs.pinggu.org/thread-3798188-1-1.html
SPSS 用什么方法分析下列数据比较合适？	http://bbs.pinggu.org/thread-3786972-1-1.html
logistic 模型的问题	http://bbs.pinggu.org/thread-3725923-1-1.html
SPSS 跑有序多分类 logistic 回归为什么没有 OR 值？	http://bbs.pinggu.org/thread-3781702-1-1.html
请教几个 logistic 回归的问题：P for trend，标准回归系数	http://bbs.pinggu.org/thread-3780537-1-1.html

3.3 纵向数据回归

问 题	链 接
如何提取特征值，建立什么模型？	http://bbs.pinggu.org/thread-3789502-1-1.html
面板 logit 模型回归中想得到经过异方差修正的标准误怎么做？	http://bbs.pinggu.org/thread-3873159-1-1.html
关于 panel data 的分析	http://bbs.pinggu.org/thread-2927230-1-1.html
如何提取特征值，建立什么模型？	http://bbs.pinggu.org/thread-3789502-1-1.html
纵向数据分析方法有哪些？	http://bbs.pinggu.org/thread-3210537-1-1.html

3.4 一级检验

问 题	链 接
SPSS 里进行回归分析时控制变量的 sig 大于 0.05，这样可以继续进行分析吗？	http://bbs.pinggu.org/thread-3754007-1-1.html
回归系数的变化趋势与 SIG 值的变化趋势	http://bbs.pinggu.org/thread-3779667-1-1.html
t 分布当中自由度，是源于什么定义的	http://bbs.pinggu.org/thread-3788442-1-1.html
只有均值，如何分析？	http://bbs.pinggu.org/thread-3788747-1-1.html

截面数据的系数有什么意义吗？	http://bbs.pinggu.org/thread-3797035-1-1.html
单侧检验与双侧（双尾）检验有什么区别？适用范围分别是什么呢？	http://ask.pinggu.org/?q-695.html
回归系数和显著性检验	http://bbs.pinggu.org/thread-3664870-1-1.html
请教：用标准误怎么解释回归系数	http://bbs.pinggu.org/thread-2209530-1-1.html

3.5 二级检验

问 题	链 接
请问怎么用 STATA 做怀特检验？	http://bbs.pinggu.org/thread-3772295-1-1.html
STATA 下 WLS 的命令问题	http://bbs.pinggu.org/thread-2364679-1-1.html
STATA 时间序列数据自相关的修正可以用 prais 语句吗	http://bbs.pinggu.org/thread-3511871-1-1.html
线性回归检测异常值使用问题	http://bbs.pinggu.org/thread-2991704-1-1.html
如何根据皮尔逊相关系数结果去掉有重复性测量嫌疑的指标	http://bbs.pinggu.org/thread-3759529-1-1.html
多个连续自变量对一个连续自变量的影响分析	http://bbs.pinggu.org/thread-3774894-1-1.html
如何用 STATA 命令消除多重共线性问题	http://bbs.pinggu.org/thread-3780222-1-1.html

共线性消除问题	http://bbs.pinggu.org/thread-3671761-1-1.html
线性回归中的共线性诊断要怎么看	http://bbs.pinggu.org/thread-3877838-1-1.html
关于多重共线性的问题	http://bbs.pinggu.org/thread-3873515-1-1.html

四、其他

4.1 问卷调查

问 题	链 接
新人求问，关于李克特量表的问题求教大家。	http://bbs.pinggu.org/thread-3788726-1-1.html
问卷量表怎么进行相关分析	http://bbs.pinggu.org/thread-3751445-1-1.html
身高体重超过问卷限定值怎么统计？	http://bbs.pinggu.org/thread-3764906-1-1.html
因变量是固定成绩，如何与李克特五分法取得的变量做相关呢？	http://bbs.pinggu.org/thread-3654780-1-1.html
信度分析克朗巴哈 a 系数为 0.7，不知这个信度系数是否可以？	http://bbs.pinggu.org/thread-3767612-1-1.html
SPSS 在做社会调研时，怎样分析问卷有效性。	http://bbs.pinggu.org/thread-3867035-1-1.html
求助！一个调查问卷中的统计分析问题，如何对比讲解前后人们观点的变化？	http://bbs.pinggu.org/thread-3876823-1-1.html
统计分析中如何处理打分题目的分析	http://bbs.pinggu.org/thread-3867755-1-1.html

关于量表设计的“降维”	http://bbs.pinggu.org/thread-3512019-1-1.html
指标体系里部分指标数据通过统计获取，部分指标数据发问卷收集，想求各指标权重，请问什么方法能够做到？	http://ask.pinggu.org/?q-609.html
问卷调查针对自变量、中间变量和因变量，问卷发放对象各为一类，对实证结果科学性有何影响	http://ask.pinggu.org/?q-651.html

4.2 相关分析

问 题	链 接
SPSS 相关性分析	http://bbs.pinggu.org/thread-3754196-1-1.html
求助 Pearson 相关系数和 Spearman 相关系数 STATA 命令分别是什么？	http://bbs.pinggu.org/thread-3790443-1-1.html
简单相关分析 三种检验方法得到的结果不一样	http://bbs.pinggu.org/thread-3764550-1-1.html
[学习资料] 典型相关分析的结构图怎么画	http://bbs.pinggu.org/thread-3750467-1-1.html
[统计软件与数据分析] SPSS 中年龄的相关分析	http://bbs.pinggu.org/thread-3834195-1-1.html

4.3 因素方差分析

问 题	链 接

求统计学大神	http://bbs.pinggu.org/thread-3749757-1-1.html
单因素方差分析方差不齐时该怎么办	http://bbs.pinggu.org/thread-3793631-1-1.html
[疑难问答] 求教多水平分析	http://bbs.pinggu.org/thread-3527139-1-1.html

4.4 非参数检验

问 题	链 接
求教，秩和检验用那个值比较好？	http://bbs.pinggu.org/thread-3771358-1-1.html
非参数检验求问	http://bbs.pinggu.org/thread-3773593-1-1.html
[疑难问答] 特殊的独立样本 T 检验怎么做	http://bbs.pinggu.org/thread-3513603-1-1.html

4.5 其他模型（灰色关联度、多层线性模型（HLM））

问 题	链 接
[疑难问答] SPSS 多元回归分层回归的一些问题	http://bbs.pinggu.org/thread-3875957-1-1.html
HLM 模型问题！	http://bbs.pinggu.org/thread-3770300-1-1.html
多层次模型中层 2 的数据可以是分年的吗？	http://bbs.pinggu.org/thread-3793181-1-1.html
灰色关联度分析	http://bbs.pinggu.org/thread-3747004-1-1.html

4.6 期刊

问 题	链 接
查找文献时，找不到这种数据的统计方法！	http://bbs.pinggu.org/thread-3785075-1-1.html
发表论文，是要先选择目标期刊呢，还是先把论文写好再找目标期刊呢	http://bbs.pinggu.org/thread-3785075-1-1.html

"长期以来，计量经济学的教学就以理论知识为主。学生对实际问题的思维和应用能力一般都被认为是天赋，而非努力培养的结果。院校招生的时候也很看重学生在掌握理论知识之后，有创造性的应用理论知识的能力。遗憾的是，很多人，即使是顶级名校毕业，在这个方面也不是非常成功。在这个背景下，这本书能够以独立于理论书籍却又不脱离理论的方式，探讨计量统计学习方法和心得，从早期开始培养学生，强调思维方式和学习态度，会对很多学生有所帮助。"

——夏目贵志
美国某高校教师
主要承担本科和MBA阶段的宏观、统计和计量经济学课程

"我一直认为一本好的计量书对于刚入门的经济学学生是多么的重要。虽然我没有从事研究工作，但当初在Angus Deaton的计量课上学到的知识让我至今受益无穷。本书作者从自己的亲身经历出发，敞开心扉，循循善诱，一定能让读者们有所收获，在经济学研究的道路上少走弯路。"

——wwqqer
普林斯顿大学金融硕士
曾在掌握全球经济命脉的纽约联储局任研究助理

"接到盛邀，通读了一遍本书，感觉整体布局合理、思路清晰、有理有据、语言朴实，对初学计量统计的人士会大有脾益，非常推荐！"

——长风神犇
北京大学光华管理学院经济学博士在读
央企直投部高级项目经理

"作者的经历也是广大计量人的的经历，这种'痛，并快乐着'的心路历程每个计量人都会深受感受。很有幸能够读到本书，吸取营养，受到启发，从此不再是一个人在战斗！"

——Jonerice
经管之家坛友

"能读到这么优秀的内容，这个中午实在是赚了！"

——明阳^_^
经管之家坛友

"九月份要开学读博了，看到这些文章很有启迪，最近自己在用stata跑伍德里奇的导论还有三章跑完了，整个过程很痛苦，但是看完这些，自己觉得这个过程是必须的，好文必须要赞。"

——zhoujielunli
经管之家坛友